Australian Government

Patent Office
Canberra

I, TESSA ROBINSON, EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2002952696 for a patent by GENOMICS RESEARCH PARTNERS PTY LTD as filed on 14 November 2002.

WITNESS my hand this
Thirtieth day of June 2009

T. Robinson

TESSA ROBINSON
EXAMINATION SUPPORT AND SALES

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

Regulation 3.2

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

for the invention entitled:

"Status Dctermination"

The invention is described in the following statement:

STATUS DETERMINATION

Background of the Invention

The present invention relates to a method and apparatus for determining the status of a subject, and in particular for determining the ability of a subject such as a horse to compete in a sporting and/or racing event.

Description of the Prior Art

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that the prior art forms part of the common general knowledge in Australia.

A condition of a performance animal, for example a racehorse, may typically be determined by conventional means such as a blood profile test and clinical appraisal. However, these tests are of limited value because a correlation between results of a blood profile test or clinical appraisal and a condition or state of a performance animal is minimal.

A blood profile test may be suitable for providing some information in relation to an animal that is clinically diseased or ill, but is rarely suitable for determining fitness to perform of an animal, particularly if the animal is healthy according to use of current clinical appraisal methods, and particularly if the animal cannot communicate information about its condition. Although blood profile tests are relatively inexpensive and easy to perform, they do not provide assessment of a wide range of conditions, correlations between test results and conditions of performance animals are poor, are limited to assessment of a few diseases, and are sometimes only useful in assessment of advanced stages of disease where clinical intervention is too late to prevent significant loss of performance.

Alternative diagnosis or assessment procedures are often complex, invasive, inconvenient, expensive, time consuming, may expose an animal to risk of injury from the procedure,

and often require transport of the animal to a diagnostic centre.

A final report of the results of a blood test to an end user, eg. a trainer, often requires involvement of multiple parties each providing separate input to the report. For example, a veterinarian may collect a blood sample, the sample is transported or sent to a laboratory for analysis, personnel in the laboratory perform an analysis using machinery on the blood sample, automated results from the analysis, with or without a veterinary pathologist interpretation, are returned to the veterinarian who then interprets the results and provides a separate report to the trainer. The process is laborious, time consuming, subject to error and interpretation bias and may or may not contain information relevant to the end user.

Bioinformatics may be used with genetic based diagnosis of an animal's health.

Currently, it is known to use genetic information in determining information regarding an individual. This can be achieved in a number of ways depending on the information that is desired.

Thus, for example, WO 01/25473 describes a method of characterising a biological condition or agent using calibrated gene expression profiles. In this case, when a subject is suspected of having a condition, a test is performed to obtain a specific profile, which is then analysed. In particular, the collected profile is compared to a predetermined profile to determine if the condition has been correctly identified. However, this suffers from drawbacks in that a preliminary diagnosis is required to allow the correct test to be performed.

US 6,287,254 describes a system that allows users to perform DNA genetic profiling to determine the susceptibility of a subject to a condition. In particular, in this example, the subject is profiled to determine the presence of predetermined genes, which in turn indicate the susceptibility of a subject to a respective condition. Again, this requires specific tests for specific conditions, and only allows the susceptibility of a subject to be determined.

Summary of the Present Invention

In a first broad form the present invention provides a method of determining the status of a subject, the method including:

a) Obtaining subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;

b) Comparing the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to distinguish each of the conditions; and,

ii) An indication of the status of the respective individual; and,

c) Determining the status of the subject in accordance with the results of the comparison.

The number of parameters is typically greater than about 100, and preferably between about 1000 and about 6000. As used herein, the term *"about"* refers to values (*e.g.*, amounts, concentrations, time etc) that vary by as much as 30%, preferably by as much as 20%, more preferably by as much as 10%, even more preferably by as much as 5%, and still even more preferably by as much as 1% to a specified or reference value.

The method typically includes determining any conditions displayed by the user.

The method can also include determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.

The method of performing the comparison usually includes:

a) Obtaining a set of templates, the set of templates representing differences between groups of individuals; and,

b) Using the templates to classify the subject data into a respective one of the groups.

The method may further include determining one or more conditions displayed by the subject in accordance with the determined group.

The parameters can be representative of the level or abundance of an agent in the subject or in a biological sample obtained from the subject. The agent may be selected from one or more of:

a) A nucleic acid molecule;
b) A proteinaceous molecule;
c) A carbohydrate;
d) A lipid;
e) A drug;
f) A chemical;
g) A gas;
h) A cell;
i) A pathogenic organism; and,
j) A non pathogenic organism.

By "*obtained from*" is meant that a sample such as, for example, a nucleic acid extract or polypeptide extract is isolated from, or derived from, a particular source. For example, the extract may be isolated directly from a tissue or a biological fluid isolated directly from the subject.

Other parameters can be measured however, such as the near IR of the subject's blood, general measurements, such as temperature, or other biological indicators.

The method usually includes:

a) Receiving confirmation of the determined status; and,
b) Updating the predetermined data in accordance with the confirmed status and the subject data.

The predetermined data can include phenotypic information of the individuals, and the subject data can include phenotypic information regarding the subject, the phenotypic information including details of one or more phenotypic traits.

In this case, the method can include comparing the subject data to predetermined data for individuals having one or more phenotypic traits in common with the subject.

In a second broad form the present invention provides apparatus for determining the status of a subject, the apparatus including a processing system adapted to:

a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;

b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to distinguish each of the conditions; and,

ii) An indication of the status of the respective individual; and,

c) Determine the status of the subject in accordance with the results of the comparison.

The apparatus can therefore be adapted to perform the method of the first broad form of the invention.

In a third broad form the present invention provides a computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of the first broad form of the invention.

In a fourth broad form the present invention provides a method of allowing a user to determine the status of a subject, the method including:

a) Receiving subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Causing the base station to:

i) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

(1) One or more parameter values for the respective individual; and,

(2) An indication of the status of each individual; and,

ii) Determine the status of the subject in accordance with the results of the comparison; and,

c) Transferring an indication of the status of the subject to the user via the communications network.

The method generally includes:

a) Having the user determine the subject data using a remote end station; and,

b) Transferring the subject data from the end station to the base station via the communications network.

The base station can include first and second processing systems, in which case the method can include:

a) Transferring the subject data to the first processing system;

b) Transferring the subject data to the second processing system; and,

c) Causing the second processing system to perform the comparison.

The method may also include:

a) Transferring the results of the comparison to the first processing system; and,

b) Causing the first processing system to determine the status of the subject.

In this case, the method preferably includes at least one of:

a) Transferring the subject data between the communications network and the first processing system through a first firewall; and,

b) Transferring the subject data between the first and the second processing systems through a second firewall.

The second processing system may be coupled to a database adapted to store the predetermined data, the method including:

a) Querying the database to obtain at least selected predetermined data from the database; and,

b) Compare the selected predetermined data to the subject data.

The second processing system can be coupled to a subject database, the method including storing the subject data in the subject database.

It is also possible to implement any one of the features of the first broad form of the invention. Thus, for example, the status may include details of any conditions of the individuals, in which case the method can include determining any conditions displayed by the user. The method may also include determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.

Similarly, the method of performing the comparison typically includes causing the second processing system to:

a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,

b) Use the templates to classify the subject data into a respective one of the groups.

The method can include having the user determine the subject data using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number. In this case, the method typically includes causing the base station to:

a) Determine the serial number from the subject data;

b) Determine a layout indicating the position of each feature on the array;

c) Determining the parameter values in accordance with the determined layout, and the subject data.

The method may also include:

a) Receiving confirmation of the determined status from the user; and,

b) Updating the predetermined data in accordance with the confirmed status and the subject data.

In this case, the features can include at least one of:

a) An oligonucleotide;

b) A peptide;

c) An antibody;

d) A carbohydrate;

e) A lipid;

f) A cell; and,

g) An organism.

The method can also include causing the base station to:

a) Determine payment information, the payment information representing the provision of payment by the user; and,

b) Perform the comparison in response to the determination of the payment information.

In a fifth broad form the present invention provides a base station for determining the status of a subject, the base station including:

a) A store method for storing predetermined data, the predetermined data including for each of a number of individuals:

i) One or more parameter values, at least one of the parameters being indicative of the current biological status of the individual;

ii) An indication of the status of the individual; and,

Best Available Copy

b) A processing system, the processing system being adapted to:

i) Receive subject data from the user via a communications network, the subject data including one or more parameter values;

ii) Compare the subject data to the predetermined data;

iii) Determine the status of the subject in accordance with the results of the comparison; and,

iv) Output an indication of the status of the subject to the user via the communications network.

The processing system can be adapted to receive subject data from a remote end station adapted to determine the subject data.

The processing system may include:

a) A first processing system adapted to:

i) Receive the subject data; and

ii) Determine the status of the subject in accordance with the results of the comparison; and,

b) A second processing system adapted to:

i) Receive the subject data from the processing system; and,

ii) Perform the comparison; and,

iii) Transfer the results to the first processing system.

The base station typically includes:

a) A first firewall for coupling the first processing system to the communications network; and,

b) A second firewall for coupling the first and the second processing systems.

The processing system can be coupled a subject database, the processing system adapted t store the subject data in the subject database.

The processing system can be adapted to:

a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,

b) Use the templates to classify the subject data into a respective one of the groups.

The method may further include determining one or more conditions displayed by the subject in accordance with the determined group.

The subject data may be determined using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number, the processing system being adapted to:

a) Determine the serial number from the subject data;

b) Determine a layout indicating the position of each feature on the array;

c) Determining the parameter values in accordance with the determined layout, and the subject data.

The processing system can be adapted to:

a) Receive confirmation of the determined ability; and,

b) Update the predetermined data in accordance with the determined ability and the subject data.

The base station of the fifth broad form of the invention may therefore be adapted to perform the method of the fourth broad form of the invention.

In a sixth broad form the present invention provides a computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of the fourth broad form of the invention.

In a seventh broad form the present invention provides an end station adapted to determine the status of a subject, the end station including a processor adapted to:

a) Determine subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Transfer the subject matter to a base station via a communications network, the base station being adapted to:

i) Compare the subject data to predetermined data for one or more individuals, the predetermined data including:

(1) One or more parameter values for the respective individual; and,

(2) An indication of the status of each individual; and,

ii) Determine the status of the subject in accordance with the results of the comparison; and,

c) Receive an indication of the status of the subject to the user via the communications network.

The end station is typically adapted to cooperate with the base station of the fifth broad form the invention to perform the method of the fourth broad form of the invention.

In a eighth broad form the present invention provides a computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to operate as an end station according to the seventh broad form of the invention.

In a ninth broad form the present invention provides a method of determining the ability of a subject to perform in a sporting and/or racing event, the method including:

a) Obtaining subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Comparing the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more parameter values for the respective individual; and,

Best Available Copy

ii) An indication of the status of each individual;

c) Determining the status of the subject in accordance with the results of the comparison; and,

d) Providing an indication of the ability in accordance with the results of the comparison.

The method of determining the status of the subject may be the method of the first or fourth broad forms of the invention.

The status of each individual typically indicates any conditions displayed by the user, in which case the method typically includes:

a) Determining any conditions displayed by the user in accordance with the results of the comparison; and,

b) Determining the ability in accordance with the determined conditions.

In a tenth broad form the present invention provides apparatus for determining the ability of a subject to perform in a sporting and/or racing event, the apparatus including a processing system adapted to:

a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Compare the subject data to predetermined data, the predetermined data includin for each of a number of individuals:

i) One or more parameter values for the respective individual; and,

ii) An indication of the status of each individual;

c) Determine the status of the subject in accordance with the results of comparison; and,

d) Provide an indication of the ability in accordance with the results of comparison.

The processing system is generally adapted to perform the method of the ninth broad form of the invention.

In an eleventh broad form the present invention provides a computer program product for determining the ability of a subject to perform in a sporting and/or racing event, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of the ninth broad form of the invention.

In a twelfth broad form the present invention provides a method of providing secure arrays for use, each array including a number of predetermined features, the method including:

a) Determining a number of respective feature layouts, each layout representing the positioning of each feature on a respective array;

b) Determining a number of serial numbers, each serial number corresponding to a respective layout;

c) Generating a number of arrays, each array being generated in accordance with a respective layout, and including the corresponding serial number thereon, the serial number being used in processing the array.

The method can be performed to provide the arrays on behalf of an entity, the method including providing an indication of the layouts and corresponding serial numbers to the entity, to thereby allow the entity to process the arrays.

The method of determining the layouts typically includes:

a) Determining a preferred layout; and,

b) Moving the position of one or more of the features from the position in the preferred layout to alternative position.

The method can include:

a) Determining the type of each feature; and,

b) Exchanging the position of one or more features having different feature types.

Brief Description of the Drawings

An example of the present invention will now be described with reference to the accompanying drawings, in which: -

Figure 1 is a schematic diagram of an example of a processing system for implementing the invention;

Figure 2 is a flow chart outlining the process implemented by the system of Figure 1;

Figure 3 is a schematic diagram of an example of a distributed architecture for implementing the invention;

Figure 4 is a schematic diagram of an example of one of the end stations of Figure 3;

Figures 5A and 5B are a flow chart of the process implemented by the system of Figure 3;

Figure 6 is a flow chart of the process for generating templates;

Figures 7A and 7B are a flow chart of the process of comparing the subject data to the templates;

Figure 8 is a schematic diagram of a second example of a distributed architecture for implementing the invention;

Figure 9 is a flow chart of the process for generating secure arrays; and,

Figure 10 is a flow chart of the process for generating subject data using the secure arrays.

Figure 11 is a flow chart of the process of data mining;

Figure 12 is a flow diagram illustrating dataflow steps in a specific example as part of a computer system capable of delivery of remote diagnostic services;

Figure 13 is a flow diagram showing steps for diagnosing a condition of an animal in accordance with a specific example;

Figure 14 is a diagram illustrating an environment for working the specific example shown in Figure 13;

Figure 15 is a flow diagram illustrating steps for preparing an array in accordance with a specific example of the invention;

Figure 16 is a flow diagram showing steps for determining a nucleic acid expression level in a biological sample; and

Figure 17 is a flow diagram illustrating steps for building a database in accordance with a specific example.

Detailed Description of the Preferred Embodiments

An example of the present invention will now be described with reference to Figure 1, which shows a processing system suitable for implementing the present invention.

In particular, Figure 1 shows a processing system 10 including a processor 20, a memory 21, an optional input/output (I/O) device 22 and an interface 23 coupled together via a bus 24. In use, the interface 23 is adapted to couple the processing system 10 to one or more databases shown generally at 11.

In use, the processing system 10 is adapted to receive subject data, which is data representative of the current biological status of a subject. The subject data is typically in the form of raw data and therefore requires interpretation to allow the status of the subject to be determined. This is achieved by having the processing system 10 compare the subject data to predetermined data stored in the database 11. The predetermined data includes data representative of the biological status of a number of individuals, together with an indication of the actual status of the individuals when the predetermined data was collected.

Accordingly, by comparison of the subject data with the predetermined data, this allows the subject data to be interpreted and the current biological status of the subject to be determined.

The manner in which this may be achieved will now be described in outline with respect to Figure 2.

In particular, at step 100 the user determines subject data in the form of parameter values representing the current biological status of the subject. In particular, the parameter values represent specific measurements of selected parameters that represent the current biological status of the subject. It will be appreciated that a number of different forms of parameters may be used, as will be described in more detail below.

At step 110 the user provides the parameter values to the processing system 10, which then operates to compare the subject data to the predetermined data at step 120. In particular, the predetermined data includes parameter values for a number of individuals having a range of different biological states.

By comparing the subject and predetermined data this allows the processing system 10 to determine the status of the subject in accordance with the results of the comparison at step 130. Thus, the processing system 10 will attempt to identify individuals having similar parameter values to the subject. The status of the subject will then be determined to be similar to that of the identified individuals.

Once the status has been determined the processing system 10 provides an indication of the status to the user at step 140.

This procedure can therefore be used to identify a wide range of conditions that may be displayed by the subject. In particular, the system can be adapted to determine the presence of one or more of a number of conditions in the subject.

In order to achieve this, each of the number of conditions must have been previously identified in the individuals. Accordingly, it is necessary to have predetermined data for a large number of individuals, with at least some of the individuals having a range of the conditions. Furthermore, it is also necessary to utilise a sufficiently large number of parameters to allow the respective conditions to be distinguished on a statistical basis.

The parameters used and typical numbers will be described in more detail below. However, it will be appreciated that the number of parameters required will generally increase depending on the number of conditions being identified.

Accordingly, it is typical for the predetermined data to include values for a large number of parameters and individuals.

It will therefore be appreciated that determination of the predetermined data is typically a time consuming and expensive procedure. It is not therefore feasible for each user wanting to implement the method to collect their own predetermined data. Accordingly, the invention is typically implemented using a distributed processing system an example of which is shown in Figure 3.

As shown, in Figure 3 the apparatus is formed from a base station 1 coupled to a number of end stations 3 via a communications network 2, and/or via a number of LANs (Local Area Networks) 4. The base station 1 is generally formed from one or more of the processing systems 10 coupled to a data store, such as the database 11, as shown.

In use, the processing system 10 operates substantially as described above to process data received via the communications networks 2, 4. The processing system 10 can then supply an indication of the determined subject status back to the respective end station 3 via the communications network 2, 4, as will be understood by a person skilled in the art.

In use, this allows the base station to be administered by an operator, that provides services allowing users of the end stations 3 to determine the status of a subject. This in turn overcomes the need for each user to obtain their own predetermined data. Furthermore, by having the base station 1 perform the comparison of the subject and predetermined data, and determine the status, this allows the operator of the base station 1 to restrict access to the predetermined data, thereby preventing the data being accessed and used by third parties. This, in turn allows the operator to charge a fee for the provision of an indication of the status of the subject, as will be described in more detail below.

In any event, it will therefore be appreciated that the system may be implemented using a number of different architectures. However, in this example the communications network 2 is preferably the Internet 2, with the LANs 4 representing private LANs, such as LANs within a company or the like.

Whilst this technique describes transferring the data electronically via the communications networks, it will also be possible to transfer data via alternative techniques such as transferring data in a hard, or printed format, as well as transferring the data electronically in a physical medium such as a floppy disk, CD-ROM or the like. Wireless transfer or the like is also possible, as will be appreciated by the person skilled in the art.

In any event, it will be appreciated that in this example, the services provided by the base station 1 are generally accessible via the Internet 2. The processing system 10 is therefore generally capable of generating web pages, or the like, that can be viewed by users of the end stations 3. Accordingly, the processing system 10 may be any suitable form of processing system that executes appropriate application software stored in the memory 21 to allow the desired functionality to be achieved. Typically however the base station 1 includes a processing system, such as a network server, web server or the like.

Similarly, the end stations 3 must be capable of communicating with the base station 1 to allow browsing of web pages, or the transfer of data in other manners. Accordingly, as shown in Figure 4, in this example, the end stations 3 are formed from a processing system including a processor 30, a memory 31, an input/output (I/O) device 32 and an interface 33 coupled together via a bus 34. The interface 33, which may be a network interface card or the like is used to couple the end station to the Internet 2 or one of the respective LANs 4.

It will therefore be appreciated that the end station 3 may be formed from any suitable processing system such as a suitably programmed PC, Internet Terminal, Lap-top, hand held PC or the like which is typically operating application software to enable web browsing or the like. Alternatively, the end station 3 may be formed from specialised hardware, such as an electronic touch sensitive screen coupled to suitable processor and memory. In addition to this, the end stations 3 may be connected to the Internet 2 or the LANs 4 via wired or wireless connections, as will be appreciated by a person skilled in the art.

Operation of the system to determine the status of the subject will now be described in

more detail with reference to Figures 5A and 5B.

In particular, as set out in Figure 5A the process begins at step 200 with the user determining the parameter values for the subject. The parameter values are then encoded as subject data by the end station 3 at step 210. This is typically achieved in accordance with a predetermined algorithm such that the subject data has a predetermined format that can be interpreted by the base station 1.

At step 220, the user accesses the base station 1 using the end station 3.

At this stage, the user of the end station 3 will typically be required to either register with the base station 1 or supply a predetermined user name and password. In particular, this is performed to allow the base station 1 to determine the identity of the user and therefore confirm that the user has authorisation to utilise the services provided by the base station 1 and/or to ensure that payment can be obtained for the provision of the services.

It will be appreciated that the user name and password will typically be provided when the user registers with the base station 1 on a first occasion. At this point the user has to make provisions for payments, such as the provision of account details, thereby allowing the operator of the base station 1 to charge the user for the services provided.

The user name and password will then be generated and subsequently verified in the normal way. Alternatively, identification of the user can be achieved in accordance with cookies stored at the end station 3, or an identifier associated with the end station 3, which may for example be the MAC (Media Access Control) address of the end station interface 33, or the like.

In any event, it will be appreciated that access to the services provided by the base station 1 is generally limited to authorised users.

In any event, when the user accesses the base station 1, this is typically achieved by

accessing respective web pages generated by the base station 1. This allows the user to select the respective services required, which in this example is an indication of the status of a subject.

Once the user has been authorised, the user will be transferred to a secure environment to allow the subject data to be transferred to the base station 1 for processing. This is typically achieved by implementing an SSL (Single Socket Layer) connection between the base station 1 and the end station 3. This provides additional security and in particular, to ensure that the subject data transferred between the base station 1 and end station 3 is retained confidential.

Confidentiality of the subject data and the determined status are important as the results are often used in determining the ability of the subject to compete in sporting and/or racing events, this information can be extremely valuable, especially to the gambling industry. It is therefore necessary to ensure the information is retained confidential at all times.

After accessing the base station 1 at step 220, the subject data is transferred to the base station 1 at step 230. At this point, the base station 1 will typically operate to review the subject data to ensure that it is genuine subject data, and that for example, the data does not disguise an attempt to hack into the base station 1 to obtain access to the predetermined data. This is typically achieved by having the base station 1 implement a firewall between the processing system 10 and the Internet 2 or LANs 4 to ensure that unwanted data is not received.

In any event, at step 240 the processing system 10 operates to determine the nature of the subject data.

Thus, it will be appreciated that the exact subject data provided and, in particular, the parameters for which values are provided may vary depending on the respective implementation. This will be described in further detail below. However, it will be appreciated that the subject data may be collected using arrays, in which case a number of

different arrays may be provided. Thus, in this case, the base station 1 will operate to determine the type of array being used, to allow the subject data to be interpreted.

At step 250 the processing system 10 selects at least some of the predetermined data in accordance with the nature of the subject data. Thus, for example, the processing system 10 will operate to select parameter values from the predetermined data for parameters corresponding to those contained in the subject data.

At step 260 the processing system 10 compares the parameter values of the subject data to the parameter values of the selected predetermined data. In particular, the processing system 10 operates to compare the parameter values to those obtained from a number of different individuals that between them have a range of different conditions. This allows the processing system 10 to determine one or more conditions displayed by the subject at step 270.

At step 280 the processing system 10 optionally determines the ability of the subject to compete in a sporting and/or racing event in accordance with the determined conditions. The processing system 10 then transfers an indication of at least the conditions to the end station 3 at step 290.

Thus, it will be appreciated that the system may be implemented in a variety of ways. Typically however the subject data is formed from phenotypic information representative of the current biological status of the subject. In one embodiment, the phenotypic information results from the expression of the genotype of the subject and is therefore typically in the form of information such as expression data, or the like. The expression data may relate to the level or abundance of an RNA molecule or a polypeptide. The RNA molecule includes, but is not restricted to, RNA transcripts such as a primary gene transcript or pre-messenger RNA (pre-mRNA), which may contain one or more introns, as well as a messenger RNA (mRNA) in which any introns of the pre-mRNA have been excised and the exons spliced together, heterogenous nuclear RNA (hnRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small cytoplasmic RNA

(scRNA), ribosomal RNA (rRNA), translational control RNA (tcRNA), transfer RNA (tRNA), eRNA, messenger-RNA-interfering complementary RNA (micRNA) or interference RNA (iRNA) and mitochondrial RNA (mtRNA). Suitable polypeptides that are contemplates by the present invention include enzymes, receptors, immunoglobulins, hormones, cytokines, chemokines, neuropeptides, adhesins, glycoproteins and the like. Alternatively, the expression data may relate to the level or abundance of a carbohydrate including monosaccharides, oligosaccharides and polysaccharides.

When the phenotypic information relates to expression data, these are typically obtained by any suitable qualitative or quantitative technique. However, where it is necessary to determine the level or abundance of a multiplicity of different expression products, it is preferable to use multiplexed analysis techniques including arrays distinctly detectable beads as is well known in the art.

The subject data may optionally contain genotypic information including genetic information carried in the chromosomes and extrachromosomally. Such data may be obtained from genetic mapping, genetic screening, pedigree, family history and heritable physical and psychological characteristics.

In this case it will be appreciated that the expression data collected may be relevant to a respective condition that is already diagnosed in the subject. However, advantageously the present invention can be utilised to detect previously undiagnosed conditions. In particular, this can be achieved by collecting sufficient parameter values and then comparing these to the predetermined data which is being collected for individuals having a range of conditions. This then allows conditions to be identified before symptoms are necessarily visible.

This can therefore be used in situations such as diagnosing conditions in animals. This is particularly advantageous as the animals are unable to provide information regarding any conditions from which they may be suffering. Thus, in the case of race horses for example, race horses can suffer from a number of conditions, such as overtraining,

respiratory illness, or the like, which can be difficult to detect. In contrast to a human athlete, who can usually communicate any symptoms to a trained medical practitioner, horses are unable to communicate to vets and therefore can only be examined passively. Accordingly, the present invention allows a vet or other medical practitioner to perform an analysis of the subject and in particular their current biological condition and determine whether the subject is suffering from any conditions.

However, it will be appreciated that this is also useful for diagnosing conditions in humans, where the human may not be aware of the condition. This is particularly the case with high performance athletes where a minor condition may not be noticeable to the athlete directly, but may have an impact on the athlete's performance.

The system is also useful for diagnosing conditions in situations where the athlete is trying to keep the condition secret, for example, in the case of drug testing to detect banned substances used by the athlete.

In order to be able to identify a significant number of conditions successfully, it is necessary to have a statistically significant quantity of predetermined data. In particular, it is necessary to have predetermined data obtained from one or more individuals suffering from a respective condition to allow the condition to be identified, and the sample size will therefore have to be sufficiently large to ensure this occurs. For example, if the chance of an individual from a general population having a specific condition is 1 in 100, it will be necessary to sample at least 100 individuals to ensure at least one individual having the condition is sampled. In fact, it would in this case be typical to sample at least 1000 individuals, to ensure that one or more individuals having the condition are identified.

Furthermore, the more data available from individuals suffering from the condition the better as this allows distinctions to be drawn between individuals suffering from different types of conditions.

The number of parameters required will depend on the number of conditions to be

distinguished. In particular, it will depend on factors such as:

- The presence and detection of unknown conditions;
- The range of conditions to be identified within the population;
- The levels of incidence of each condition in the population;
- The ability to distinguish between the conditions.

It will be appreciated that as individuals, including performance animals such as race horses, can suffer from a wide variety of conditions, then it is preferable for a large number of parameters such as 3,000 to 5,000 to be used. However, this number can be significantly lower if only a minor number of conditions are to be identified. Thus for example, the number of parameters used may be anywhere from 10 up to 10,000, or more. Suitably, the number of parameters employed are at least about 20, preferably at least about 50, more preferably at least about 100, even more preferably at least about 150, even more preferably at least about 200, even more preferably at least about 300, even more preferably at least about 500, even more preferably at least about 1000, even more preferably at least about 1500, even more preferably at least about 2000, even more preferably at least about 4000, even more preferably at least about 6000, even more preferably at least about 8000, and still even more preferably at least about 10000.

In addition to this, the effect of a condition on an individual may also vary in accordance with additional phenotypic information relating to a particular characteristic or set of characteristics of the subject, as determined by interaction of the subject's genotype with the environment in which it exists. In this embodiment, such 'characteristic data' may be selected from age, sex, height, length, weight, ethnicity, race, breed of animal, feeding patterns, exercise patterns, medication supplied, nutritional or growth supplements supplied, nutritional analysis, hair colour, skin colour, eye colour, body composition, fat composition, water retention, obesity, transcriptomic profile, proteomic profile, blood type, tissue type, endocrine function, immunological function including cellular and humoral immune function, tolerance, allergy, transplant rejection, cancer, hyperplasia, gastrointestinal function, neurological function, kidney function, heart function, brain function, pancreatic function, bone function, joint function, sexual or reproductive

function, metabolic load, toxicological profile, substance abuse including drug dependency, inborn errors of metabolism, infectious disease including viral infection, bacterial infection, mycobacterium infection, parasitic infection, prion function, prosthesis, tissue reconstruction, surgery, pain and the like. The phenotypic information may also include demographic information, which can be important for monitoring the spread of a condition globally, as well as to allow analysis to take account of conditions that are limited to predetermined areas. Thus, it is generally preferable to additionally collect characteristic data together with the expression data for the individuals.

It will be appreciated that in view of the amount of predetermined data involved, it is not generally feasible to compare subject data determined for any one subject to all the predetermined data. Accordingly, some pre-processing of the predetermined data is generally performed. This process is generally known as data mining and will be described in more detail below and will now be outlined with reference to Figure 6.

In particular, at step 300 the operators of the base station 1 operate to collect the predetermined data, including genotypic and phenotypic information, from a number of individuals.

In this example, it will be assumed that the parameter values that form the predetermined data correspond to expression data and in particular, concentration quantities, abundances, or ratios of respective expression products obtained from an array or the like. The phenotypic information will typically be provided from a study of the respective individual, and is preferably provided in a standard format to allow the information to be correctly interpreted by the base station 1.

At step 310 the processing system 10 operates to generate parameter vectors for each individual. Each parameter vector is formed from a vector containing each of the parameter values at a respective location within the parameter vector. The processing system 10 then operates to consider the position of the parameter vectors in an N-dimensional space, where N corresponds to the number of parameters at step 320.

At step 330, the processing system 10 operates to determine groups of the individuals in the predetermined data.. In particular, the groups may be identified by considering individuals having similar physical conditions.

Thus, the processing system 10 will operate at step 340 to produce diagnostic templates that are used to characterise the groups of individuals identified in the predetermined data, and to classify individuals into these groups. It will be appreciated that there is a multiplicity of ways of defining such diagnostic templates for example regularised discriminant analysis, Support Vector Machines, recursive partitioning, artificial neural networks, or the like, as will be described in more detail below with respect to data mining.

Having identified the groups, and the templates, the processing system 10 will operate at step 350 to characterise the ability of the diagnostic templates to predict group membership, by applying the templates to the individuals comprising the predetermined data. This characterisation and validation may be achieved, for example, by k fold crossvalidation, and the construction of permutation distributions.

Once the templates have been generated, these are then stored in the database at step 360, together with any other summary statistics necessary to provide statistically efficient prediction using the template.

It will be appreciated that the definition of groups may be done for respective individuals irrespective of the phenotypic traits. Thus, for example, if all individuals suffering from a condition tend to have similar parameter values, then all the individuals having the condition will be contained in the same group irrespective of each individual's phenotypic traits.

However, if different phenotypic types have distinct parameter values for the same condition, then a respective template will be defined for each phenotypic group. Thus for example, a template may be defined for male horses having a respiratory condition, with a

separate template being defined for female horses having the same respiratory condition.

In addition to this, at least one group will correspond to healthy individuals not having any conditions. Thus a template will be generated corresponding to a healthy individual. It will be appreciated that this can be used in determining if an individual has an unidentified condition, as will be described in more detail below. This can also be used to identify sub-clinical diseases or conditions that are not previously apparent through existing diagnostic techniques.

Once the templates have been generated, it is then possible to operate to compare the subject data to the predetermined data. The manner in which the comparison is performed will now be described with reference to the flow chart shown in Figures 7A and 7B.

In particular, at step 400, the user determines parameter values, and phenotypic information relating to the subject. At step 410, the end station 3 is used to generate subject data in accordance with the determined parameter values and phenotypic information. At step 420 the user transfers the subject data to the processing system 10 as described above.

At step 430, the processing system 10 extracts the parameter values and the phenotypic data from the subject data, and then uses the parameter values to generate a parameter vector at step 440.

The processing system obtains one or more of the templates from the database at step 450. At this point, it will be appreciated that the templates may be selected in accordance with the phenotypic information, such that the subject parameter vector is only compared to templates having suitable phenotypic traits. Thus, for example, it will be appreciated that if the subject is a male horse, then it may be pointless comparing the subject parameter vector to a template representing a group of female horses having a respiratory disease.

However, if a template corresponds to a group of individuals having a range of phenotypic

traits, then this template will be used to predict group membership using the subject parameter vector, at step 460. It will be appreciated that there is a multiplicity of ways of predicting group membership from the subject parameter vector, just as there is a multiplicity of ways of constructing group templates, as will be appreciated by persons skilled in the art.

At step 470 the processing system 10 operates to determine the uncertainties in group prediction using the subject parameter vector and templates in the N dimensional vector space. These uncertainties are expressed as probabilities that the test subject has a condition previously characterised by membership of one of the groups in the predetermined data.

It is apparent to those skilled in the art that there is a multiplicity of ways of constructing these uncertainties, each appropriate for a different method of template construction and group prediction. For example uncertainties may be based on some measure of distance between the subject parameter vector and a group template, or by a Bayes rule applied to a set of discriminant functions.

It will be appreciated therefore that the templates may be based on specific values such that they represent a single point in the N dimension vector space. Alternatively however the templates may correspond to ranges such that each template defines a range of parameter values for which the subject would have the respective condition. Thus, this effectively defines decision boundaries in the N dimensional space, such that if the subject parameter vector falls within the decision boundary, this indicates that the subject has the respective condition.

If the parameter vector is approximately equidistant to two or more templates, this may indicate that there is a chance that the individual either has a previously undetermined condition, or alternatively is suffering for example from a combination of the two conditions. It will be appreciated that templates may be generated for common combinations of conditions, as well as single conditions.

Finally, it will be appreciated that the presence of the template for healthy individuals allows a healthy subject to be determined. If the subject parameter vector is significantly separated from this template, this will indicate that the subject is generally unhealthy, and this allows previously unidentified conditions to be determined, for example, if the subject parameter vector is not near any of the other templates.

It will also be appreciated that the magnitude of the parameter values will allow the severity of conditions to be determined. Thus, for example, the greater a difference in magnitude between the parameter values for a healthy subject compared to a subject suffering from a condition will generally indicate a greater severity of the respective condition.

Similarly, it will be appreciated that groups may be defined for different severity of condition. Thus, for example, a first group may be defined for the initial stages of a condition that is treatable, whilst a second group is defined for the same condition when it has progressed beyond the initial stages and is no longer treatable.

Finally, a direct comparison of the subject parameter values can be made with the predetermined data for other individuals suffering from the same condition, can also be used to allow the severity of the conditions to be determined.

In any event, at step 480, the processing system 10 interprets the separation of the parameter vector from the templates and uses this to determine any conditions displayed by the subject. An indication of this is then transferred to the end station 3 at step 490.

A number of alternatives can be implemented in the present invention.

Multiple Firewall

In particular, in the above described example it will be appreciated that users of the end stations 3 are unable to access any of the data stored in the database 11. This is performed

to ensure that the data can be retained as confidential by the operator of the base station 1.

This in turn allows the operator of the base station 1 to continue to provide indications of subject status without running the risk of users of the system obtaining the raw data stored in the database 11 and using this for their own purpose. This ensures that the operators business of providing an indication of the status for a fee is protected.

However, it will be appreciated that the security provided by the above system is in some extent limited. In particular, there is the opportunity that hacking may occur in which users of the end stations 3 attempt to infiltrate the processing system 10 and cause the processing system 10 to download data, such as the templates, from the database 11.

In order to overcome this, the base station 1 can implement a dual processing system set up as shown for example in Figure 8. In this example, the base station 1 includes a processing system 12 coupled to the LANs 4 and the Internet 2 via a first firewall 13, and a second database 14 coupled to the first processing system 12 via a second firewall 15.

In this example, the processing systems 12, 14 will be substantially similar to the processing system 10 described above, and will not therefore be described in further detail.

In use, communication with the end stations 3, including the receipt of the subject data, and provision of results, is achieved using the processing system 12. In the case of receiving of subject data, or any other requests, the received submission is analysed by the processing system 12, and any relevant information extracted. The extracted information, which is determined by the processing system 12 to be a genuine submission, can then be transferred to the processing system 14.

Thus, the processing system 12 can receive the subject data, and operate to extract the parameter values therefrom. The processing system 12 then generates the parameter vector, or the like, which is transferred to the processing system 14 for subsequent comparison with the predetermined data.

Once the comparison has been performed, the processing system 14 can determine those conditions suffered by the subject and then transfer an indication of this back to the processing system 12 through the firewall 15. The processing system 12 can then transfer an indication of this indication to the end station 3.

It will be appreciated that in this example even assuming the user is able to infiltrate the first firewall 13, the user will only be able to access previously submitted requests and the results determined therefrom. The presence of the dual firewall system therefore makes it virtually impossible for the user to infiltrate the processing system 14 and obtain access to the data stored in the database 11.

In the remainder of the description, it will be appreciated that the processing systems 10; 12, 14 are effectively interchangeable.

Parameter Ranges

A further alternative to the present invention is for the comparison to be performed on the basis of parameter ranges defined for different conditions.

Thus for example, each condition may have associated therewith a sequence of parameter value ranges determined based on ranges of parameter values for individuals diagnosed with the respective condition. The parameter value ranges can then indicate for a respective condition the parameter values that can be expected, allowing the determined parameter values to be compared to the respective range for each condition to determine if the parameter values provided fall within a respective range.

Thus, a respective parameter range can be determined for each condition, with the parameter values determined for a subject being compared to each range, to determine those ranges within which the subject data falls.

An indication of the likelihood of the subject having a respective condition can then be

C:\WINNT\Profiles\VFA\Desktop\Gene...\nit_as revision_final.doc 14/11/02

determined statistically based on the number of individuals having the respective condition.

Multi-Level Analysis

In the example described above, a number of conditions have been defined for the respective type of individual. However, it will be appreciated that sometimes it is desirable to perform tests to focus on specific conditions. Thus for example, in the case in which a horse has an existing condition, it is sometimes desirable to monitor the development of the condition for the respective subject.

In this case, as the condition has been determined, it will not usually be necessary to consider all of the parameter values each time the analysis is performed. In particular, as a large number of parameters are provided to allow the different conditions to be distinguished, a large number of parameters will typically not be representative of the progress of a specific condition.

Thus, it is usually possible to identify a number of key parameters that are relevant to respective conditions. Thus for example, conditions relating to respective respiratory illnesses may be uniquely identified using a smaller number such as 50 parameters. In this instance, if the user is only interested in examining for the progress of this respective condition, the user can simply supply an indication of the values for the respective 50 parameters.

In this example, the processing system 10 would operate to compare the determined parameter values against parameter values of horses suffering from the condition and horses not suffering from the condition. In this situation, the manner in which the collection of the parameter values is performed may very.

In the examples described above it has been mentioned that the parameter values may include for example expression data collected using an array, for example. If the arrays are to collect values corresponding to 5000 parameters it is typical for an array to be provided

with 5000 features thereon with each feature corresponding to a respective parameter. Alternatively, 10,000 features may be provided with two features corresponding to each parameter. In any event, a person skilled in the art will appreciate that a number of variations on this are possible.

However, if only 50 parameters are to be measured, it is then possible to provide an array having 5,000 features with 100 features being used to determine the value for each parameter. This allows the parameter values to be determined far more accurately allowing a more accurate representation of the condition to be determined. In particular, more accurate comparison of the subject data with the predetermined data can be performed.

Thus, a typical sequence of events may be for a user to submit a general test having a large number of parameters similar to that described above which allows respective conditions to be first identified. Once a condition has been identified, the user can then purchase specifically designed array plates adapted to monitor the specific condition. Measurements of the parameter values relevant to the condition can then be made far more accurately allowing the progress of the condition to be monitored in detail. This can allow users to be provided with information concerning whether conditions are improving or not.

Longitudinal Analysis

In the methods described above, the subject data for a respective subject is compared to predetermined data for a number of different individuals. However, in addition to, or alternatively to this, longitudinal analysis can also be performed. In this instance, the subject data is compared to subject data previously collected for the same subject. Thus, this allows the progression of a condition within a subject to be monitored.

Again, it will be appreciated that if this is performed with a limited number of parameters as described in the multi-level analysis described above, then this allows an accurate assessment of the progression of a condition to be made.

By storing the results determined for a respective subject in the database 11, for a predetermined time period, this can allow the progression of the disease over a time period to be monitored and displayed to the user. Thus, the most recently obtained subject data is compared to earlier subject data for the same subject (and optionally predetermined ata), to determine disease progression.

Thus, for example, levels of respective parameter values can be used to indicate the severity of the disease. This can be achieved by comparing the subject data to predetermined data in the manner described above, or alternatively using other techniques. As the parameter values vary over time, this can be used to provide an indication of whether the condition is improving or worsening. This is turn can be used to monitor the effectiveness of any treatment given to the subject.

Thus for example, if it is determined that a horse has been overtrained then the obvious solution to this problem is to reduce training for a predetermined time period, or resting the horse. However, trainers will generally not want to reduce the training too much as the horse will become unfit. Similarly, worse problems can arise if the trainer resumes training too early.

Thus, in this case, the trainer can submit subject data on a periodic basis such as every week allowing the fitness of the horse to be determined on a weekly basis. An indication of this can then be transferred back to the user allowing the trainer to determine when training of the horse should resume, or how hard training should be.

This therefore allows the severity of the condition within the subject to be monitored.

Secure Arrays

Arrays, or genechip technology determines whether genes are turned "on" or "off" under certain conditions. A single gene is a stretch of DNA found in the nucleus of a cell that encodes, through mRNA, the information required to produce a particular protein. Genechips measure the level of mRNA in cells and can be directly correlated to the

amount of protein being produced in a cell.

In the example of collect subject data for horses, a horse array and a blood sample are needed. The array has DNA dotted onto its surface (DNA of the genes in horse blood cells). The DNA on the array consists of one strand of the double-stranded DNA molecule – the other strand is provided by the blood sample and is labelled with a dye.

Two strands of similar DNA will only bind to each another (hybridise) if they match in sequence. An array reader can determine the amount of mRNA in a sample (gene turned "on" or "off") by determining the amount of dyclabelled DNA that hybridises to an array.

The reader produces a value compared to a reference for every single gene on the array. The 5,000 to 10,000 values can then be compared to Genetraks' database. Genes turned "on" or "off", individually or in patterns, can then be identified and correlated to the specific conditions of a racehorse.

Various conditions in racehorses will alter the metabolism of the white blood cells, which can then be detected using the genechip technology. For example, the gene for manganese superoxide dismutase (MnSOD) may be turned "off" in respiratory inflammation. Similarly, IFNg, IL-4 may also be turned "off", and the genes for Gro1a, IL-8, TNF and MIF may be turned "on". This pattern of "gene expression" can be correlated to a specific condition, such as respiratory inflammation caused by a virus. Patterns of gene expression change as a horse succumbs to or recovers from a viral infection. As the technology and database develops, predictions on the stage of infection or influence of treatments can be made.

As described above, it is preferably to ensure that the predetermined data is retained as confidential.

However, if arrays are used in the collection of data, it will be appreciated that it would be possible to purchase a quantity of arrays and perform data mining of data obtained from

Best Available Copy

used arrays to determine new predetermined data. Thus, there is a danger that competing companies will use the arrays provided on behalf of the operator for their own purposes.

In order to be able to do this, it will be necessary for the competing entities to be able to interpret the data provided by the arrays. However, this can be overcome by utilising secure arrays. In particular, secure arrays utilise a randomisation of the layout of the array to avoid the problems of reverse engineering or the like.

The manner in which this may be achieved will now be described with reference to Figures 9 and 10.

In particular, at step 600 the operators of the base station 1 will determine a number of features to be included on the array, and provide an indication of these features to the array supplier at step 610.

At step 620, the array supplier will operate to generate a preferred array layout using a processing system. This is performed in accordance with normal operating procedures. In particular, the array suppliers will generally utilise applications software to determine a preferred array layout which optimises the array build process. The layout is generally organised so that creation of the array is simplified.

At step 630, the array supplier will operate to generate a number of randomised arr layouts. The randomised array layouts have one or more of the features positioned in alternative location when compared to the preferred array layout. In particular, the ar supplier will generally operate to move or swap the locations of one or more of the feat on the array. In order to swap features, it must be ensured that the features are of diff types.

At step 640 the array supplier will also operate to generate a corresponding num serial numbers. In particular, a respective serial number is provided for each rand array layout that is generated.

At step 650 the array supplier will operate to generate arrays in accordance with the randomised array layouts and the serial numbers. In particular, each generated array will have features positioned thereon in accordance with a respective one of the randomised layouts, together with an indication of the corresponding serial number.

It is typical for the array supplier to produce the arrays in batches with up to 1,000 arrays in each batch, with each batch being created in accordance with a difference randomised layout.

At step 660 the randomised arrays are transferred to the users for subsequent use in generating the subject data, whilst at step 570 the serial numbers, together with corresponding layouts are transferred to the base station 1.

The use of randomised arrays will slightly complicate the production process but will vastly increase the security of the arrays. In particular, third parties will be unable to utilise the arrays, as the location of features alter, which will cause the third parties to obtain varying results on different arrays, for the same sample.

Operation of the system to use the randomised arrays will now be described with reference to Figure 10. In particular, at step 700 the user will obtain a biological sample from the subject and then perform an assay process using the array at step 710.

At step 720 the user uses the end station 3 to encode the values obtained from the array as subject data, together with a serial number indication. The subject data is then transferred to the base station 1, in the manner described above, at step 730.

The processing system 10 operates to determine the serial number from the subject data at step 740. The serial number is then used to access the respective array layout stored in the database 11 at step 750.

The array layout will then be used by the processing system to interpret the subject data, and in particular, to determine the respective feature to which each value corresponds. This allows the processing system 10 to hence determine the parameter values for the respective subject data.

Operation of the invention will then be substantially as previously described above.

It will be appreciated that the serial number may also be used to check the user is an authorised user. In particular, if each user is provided with arrays having a respective serial number (a range of serial numbers), then having the array supplier provide an indication of the user and the serial number(s) to the operator, this allows the operator to verify the identity of the user. This provides an audit trail for the arrays.

Feedback

A further way in which the present invention may be utilised is to provide feedback on the accuracy of provided results.

In particular, if the base station 1 is used to provide an indication of one or suspected conditions in a subject, the user can be requested to provide an indication whether the diagnosis provided by the base station 1 is correct. This may form a requirement, such that a user will only be provided with services by the base station if they agree to this term.

In any event, the correctness of the assessment by the base station 1 can usually be determined by either treating the subject and determining if the treatment is successful, or by monitoring the development of the condition over a predetermined time period. Once it has been determined that the diagnosis is correct or incorrect, an indication of this can be transferred to the base station 1.

At this point, the respective subject data collected for the respective subject can be saved as predetermined data in the database 11, with the confirmation of the condition being used as the indication of the condition in the predetermined data.

In order to achieve this, the processing system 10 is typically coupled to a sample database that is used to store the subject data obtained from each subject. Once confirmation of the conditions is received the subject data and the condition indication is transferred to the predetermined data stored in the database 11.

It will be appreciated that this checking of the conditions is not essential to the present invention as typically the data alone will be useful. However, checking of the condition will be useful in determining the accuracy of the templates.

It will be appreciated that as further data is collected over through the feedback technique or through the use of alternative data collection methods the templates or other data can be updated allowing more accurate condition analysis to be performed.

Users

It will be appreciated that any individual may use the system. Initially at least however, it is necessary for the user to be able to generate the subject data. In the case in which arrays are used, for example, this requires the user to first collect biological material, such as blood, and then analyse the material using the array. This is generally difficult and requires skilled operators using existing technology. Accordingly, the user may have to be a skilled technician. However, it is envisaged that collection techniques will become simpler, allowing the process to be implemented by any user.

In the case of sporting or racing events, for example, the users could include:

- Athletes;
- Trainers;
- Drug testing committees (such as Olympic Officials);
- Medical practitioners (such as vets or doctors);
- Event organisers;
- Pathology labs (that would typically perform the work on behalf of an individual, such as a horse owner).

However, this is not intended to be limiting.

Reports

In general, the indication of any conditions suffered by the user, together with information concerning the ability of the subject to compete in events, or the like, is provided in the form of a report.

It will be appreciated by those skilled in the art, that the content of the report may need to be tailored depending on the type of user. Thus, for example, a trainer will not be interested in knowing about parameter values for their horse, but will rather want to know what conditions the horse has, and the severity. In contrast, if the user is a skilled medical practitioner, then there may be some benefit in having more detailed information provided thereon.

Accordingly, the processing system 10 can be adapted to generate tailored reports in accordance with report templates stored in the database 11, or the memory 21. In this case, the processing system 10 will determine the type of user, and then access a respective report template. The template will specify the type of information to be provided to the user, allowing the processing system 10 to populate the report in accordance with the results of the above described analysis.

Thus, for example, in the case of the user being a trainer, the processing system 10 can access a user report template, which will include a number of fields. The processing system will determine from the field the information required, and populate the fields accordingly. This may require some additional processing to place the information in the required form. The information will also be directed to a level the user can understand, and will therefore typically avoid the use of technical terms (such as medical terms) for non-technical users.

Thus, the processing system may be adapted to determine the condition and severity. This

is then used to access a look-up table, which indicates how serious the condition is to the subject. Thus, the LUT may indicate that the condition is serious and medical condition should be obtained. In this case, the report may therefore indicate merely that the subject has a condition and medical attention should be obtained. It will be realised that the advice may depend on phenotypic data. Thus, a young horse may be more or less likely to require medical treatment for a given condition that an older horse.

For skilled medical practitioners however, more detail may be required, in which case, the processing system may be adapted to indicate not only the condition and severity, but also provide an indication of various important parameter values (such as red or white blood cell counts), to allow the medical practitioner to determine what action to take.

It will be appreciated that the information displayed may depend not only on the user, but also the respective condition. Furthermore, the information could be displayed graphically or as numerical or textual information.

As the completion of the report template is automated, it will be appreciated that users may be allowed to submit their own report templates, in accordance with predetermined criteria, allowing the user to have reports generated in their desired format.

Finally, the processing system 10 can be adapted to provide other advice. This can include for example, recommendations for changes in feeding habits, or the like. In general medical advice would not be given due to the issue of liability. However, it will be appreciated that the operator of the base station 1 could provide a medically trained individual to provide medical advice if required.

Architecture

A range of different architectures may be implemented in addition to those described above. Whilst these will not be described in detail, it will be appreciated that any form of architecture suitable for implementing the invention may be used. However, one beneficial technique is the use of distributed architectures. In particular, a number of base

stations 1 may be provided at respective geographical locations. This can increase the efficiency of the system by reducing data bandwidth costs and requirements, as well as ensuring that if one base station becomes congested or a fault occurs, other base stations 1 could take over. This also allows load sharing or the like, to ensure access to the system is available at all times.

In this case, it would be necessary to ensure that each database 11 contains the same information and templates such that the use of different ones of the base stations 1 would be transparent to the user.

Subject Data

The subject data may be selected from any expression product of the genome or characteristic or set of characteristics of the subject whose levels or abundance may vary within the subject or between two or more different subjects depending on their status. The data include, but are not restricted to, biological, physiological and pathological data of the subject. Examples of biological data include, transcriptomic profiles, proteomic profiles, enzyme function, receptor function, and the like. Physiological data may be selected from age, sex, height, length, weight, ethnicity, race, breed of animal, feeding patterns, exercise patterns, medication supplied, nutritional or growth supplements supplied, hair, skin and eye colour, fat composition, obesity, blood type, tissue type, endocrine function, immunological function, gastrointestinal function, neurological function, kidney function, heart function, brain function, pancreatic function, bone function, joint function, prosthesis, tissue reconstruction, surgery, pain and the like. Examples of pathological data include infectious disease including viral infection, bacterial infection, mycobacterium infection, parasitic infection, prion function, cancer, transplant rejection, inflammatory diseases such arthritis and fibrosis, toxicological profiles, substance abuse including drug dependency and the like.

Data Mining

Data Mining

C:\WINNT\Profiles\VFAU\Desktop\Documents and Settings_Local.doc 14/11/02

The system uses a self learning classification system, in which diagnosis is made using a historical database of test results (the predetermined data), which is updated as each test sample (subject data) is recorded. The historical database is typically maintained on a server.

Clinical application of the system can be used to diagnose a subject such as an animal with an unknown clinical or performance state. That is, the animal may or may not have some disease, or may or may not be race-ready. A metabolic profile is measured for the animal subject.

In a preferred example, the metabolic profile is comprised of expression signatures measured on an oligonucleotide chip. In a preferred example the metabolic profile is compared with a set of pre-computed diagnostic signatures (templates), and together these are used to predict the health status of the subject. In a preferred example, prediction will include probabilistic estimates of uncertainty, and be accompanied by a list of possible differential diagnoses.

Diagnostic signatures are computed by data mining a historical database, which contains metabolic profile data on subject animals (predetermined data), and associated clinical information on subject health and performance status. These historical data use the same metabolic profile measurement technique as is used in clinical application. In a preferred example, these metabolic profiles are comprised of expression signatures measured on an oligonucleotide chip.

Data mining may be performed using a number of techniques including:

- Regularised discriminant analysis for high dimensional data, as described by Kiiveri (1992) Canonical variate analysis of high dimensional spectral data. Technometrics 34 pp. 321-331.
- Diagonal discriminant analysis as described by S. Dudoit, J. Fridlyand, and T. P. Speed (2002). Comparison of discrimination methods for the classification of

tumors using gene expression data. Journal of the American Statistical Association, 97 (457), pp. 77—87.

- Support Vector Machines as described by M P. S. Brown, W Noble Grundy, D Lin, N Cristianini, C Sugnet, T S. Furey, M Ares, Jr., D Haussler (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Science. 97(1):262-267. and Y. Lee, Y. Lin, and G. Wahba (2002) Multicategory Support Vector Machines, Theory, and Application to the Classification of Microarray Data and Satellite Radiance Data. Technical Report 1064. Department of Statistics, University of Wisconsin-Madison.
- Bayesian kernel fitting algorithms.
- Tree based recursive partitioning Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984) Classication, augmented by Bagging Breiman, L. (1996) Bagging predictions, Machine Learning 26(2) pp. 123-140 and Boosting Breiman, L.(1998) Arcing classifiers. Annals of Statistics 26(3) pp. 801-849

It will be apparent, to practitioners skilled in the art, that other data mining procedures may be used to replace those identified above, without materially changing the nature of the invention.

Diagnostic signatures are combined with test subject metabolic profiles to produce a diagnosis. In one example, (where data mining was based on regularised or diagonal discriminant analysis), prediction is based on a Bayes classification rule, and estimates of uncertainty are based on posterior probabilities of class membership.

In another example, (where data mining is based on Support Vector machines), classification is based on the support vectors, and uncertainties are estimated from distance of the test profile to the decision boundary. In another example (where data mining is based on recursive partitioning) classification is based on the estimated decision tree, or averaged over multiple decision trees.

C:\WINNT\Profiles\VPAU\Desktop\Genetic Infrastructure_ver revision_GeneLogic-14-11.doc

It will usually be the case that even for an animal with an unknown clinical condition or performance status, some clinical or performance conditions are known. For example, it may not be known whether or not the animal has disease A, but it *is* known that the animal has disease B and does not have disease C. When test samples are recorded, the historical database is updated to include the test sample, and any known concomitant clinical or performance information.

It will usually be the case that an animal is tested more than once during a period of investigation. Re-testing may occur at a time when an earlier unknown clinical condition has become known. For the example given above, it may be the case that at a time of re-testing for race-readiness it is known that during the initial test the animal *did* have disease A. Provision is made to allow updates to and modification of the clinical data obtained for each test subject, as diagnosis is confirmed or modified.

In one example, data mining is repeated at regular intervals as the historic database grows. Test records added to the historic database will frequently contain only partial clinical or performance data. For any given clinical or performance factor, data will be filtered to remove subjects for which the particular characteristic is unrecorded. The data mining algorithm will then be used to construct new diagnostic signatures for the given clinical or performance characteristic. The procedure of filtering and mining is repeated for each characteristic of interest. In this way, the sample sizes used to obtain diagnostic signatures are constantly increasing, and predictive performance improves. The system becomes self-learning.

The overall process is illustrated by Figure 11: which shows the flow of information and processing in the self-learning diagnostic system.

It is apparent that the Historical database must be initialised, and preliminary data mining conducted before clinical application of the diagnostic system. The database will be initialised using a training set comprising data from animals with known metabolic conditions. Appropriate experimental design is vital to the construction of the initial

training data set. Empirical predictors derived using data mining are susceptible to artefactual relationships, involving nuisance factors – such as regional differences in diet and husbandry. For this reason, the training data set must be obtained from a multicentre trial, and stratified appropriately.

Specific Examples

Specific examples are set out in more detail in Appendix A. These are for the purpose of demonstration only and are not considered to be limiting.

Persons skilled in the art will appreciate that numerous variations and modifications will become apparent. All such variations and modifications which become apparent to persons skilled in the art, should be considered to fall within the spirit and scope that the invention broadly appearing before described.

Thus, for example, the above description has focussed on the testing of a general subject. It will be appreciated that this is most advantageously used for performance animals to identify conditions that may lead to a decrease in performance. This allows trainers to identify problems with horses or other animals before they would be noticeable using existing techniques. This is of particular benefit in the horse racing industry as it allows problems to be identified in advance, which can in turn allow the conditions to be corrected before they effect the horses performance, which in turn can result in a vast loss of earnings for the trainers and owners of the horse.

However, the technique may also be applied to any subjects, including humans.

It will be appreciated that different predetermined data will be required for each type of subject being assessed.

APPENDIX A

Figure 12 is a flow diagram illustrating one specific example of an information technology architecture and data flow as part of a remote delivery service process. External users are shown as Class One 505, Class Two 510, and Class Three 515 that are interested in obtaining information regarding their respective gene expression results when using the proprietary gene expression analysis service. These users may include, for example, pathology laboratories, drug laboratories, pharmaceutical companies, collaborators, medical and/or veterinary practitioners or similar, owners of performance animals, athletes and/or athletic trainers. Each of these users 505, 510, 515 will be interested in different aspects of the gene expression results and will therefore interact in a different fashion, but all will interact remotely via an user interface module 520.

Interface 520 may, for example, be a browser-based interface as found on most computers and delivered via web pages on the world-wide-web (the Internet). The initial interaction to the user interface module 520 will be via a controlled firewall and web server. The firewall will be the first line of defence against unwanted and unauthorised intrusion. Port blocking techniques and protocol restrictions will be imposed at the firewall. The firewall and web server environment will be fully maintained with the latest security patches to ensure currency of protection against hackers and intrusion. Each user will establish a secure connection 525 (user authentication and establish secure web connection) to ensure confidential identification in both directions for the user and service delivery provider. The security is managed by a customer access management system 565 that controls access of users 505, 510, 515. Such security measures are commonly used in the art and one embodiment would be use of SSL (secure socket layer) technology and digital signatures. Further security layers can be added at this interface if required and might include challenge/response component such as continuously changing numerical keys in possession of the user and available in plastic card format and trusted networks.

Class One and Two Users 505, 510 are shown sending information as a query 530 and 531, that includes a question regarding health or condition status of an animal (interpretation

C:\WORD\Prosecution\VPA\Desktop\Documents\Specification_as_modified_final.doc-14/11/02

request), sample details, gene expression results, clinical information, pathology laboratory results, gene identities, gene sequences, collaborative requests, etc. Class Three Users 515 are shown sending information 535 as a query including interrogation requests regarding a health status of individual animals/athletes or groups of individual animals/athletes.

Queries 530 and 531 may contain formatted gene expression and clinical information as a request, one such embodiment would employ the use of digitally signed XML documents to ensure authenticity and content of the request. Other authentication, authorization and encryption and key management standards will be applied as they become available.

As a further security measure to protect central databases 590, from outside unauthorised access, queries are temporarily stored in a transaction staging module 540 and queries 532 and 533 will be drawn into respective pathology service module 550 and collaborative services modules 555 only on request from the service module. This process may employ a second firewall and may be configured to further restrict network traffic. This firewall will only permit internal requests from 550 555 560 to pass through the firewall. All other network traffic will be blocked as will unnecessary ports and protocols. Respective pathology services module 550 and collaborative services module 555 include special software capable of servicing requirements of the different types of users 505, 510. Pathology services module 550 and collaborative services module 555 are shown in communication with each other. Core central databases 590 store genetic information (genetic database) 591, sample and gene expression information (sample database) 593, and correlative data (correlative database & heuristics) 595. The genetic information stored in genetic database 591 is used to create gene expression devices Design details 592are also stored in the sample database which contains gene location information on the device and are used to interpret results from such a device.

The genetic database 591 is also used to provide gene identification and gene sequence information to collaborative services module 555 and collaborative services 575 (eg. interpretations, gene lists and gene sequences) to Class Two users 510. Information in the sample database 593 can be clustered together based on similarity using computer

algorithms such as K-means, principal component analysis (PCA) and self-organising maps, commonly available in packages provided by companies such as spotfire, silicon genetics, and at higher levels of interpretation, Omniviz. These clusters amount to identified correlations 594 between gene expression and sample information and are stored in various formats, in the correlative database 595. An heuristic or neural network or rule-based computer software system pre-programmed with rules or training sets takes queries 534 (eg. expression details and sample details), stores these details in the sample database 593 and then compares the query pattern to those already stored in the correlative database 595 and produces standardized reports and correlation details 570 (according to the rules of the heuristic program). Correlation details are converted to useful information such as gene expression correlation results, for example a fully formatted report to include interpretations 571 and interpretations 575 (and optionally genes lists and gene sequences) and are securely delivered back to the requestor via the internet to Class One and Two users 505, 510.

Financials database 597 keeps track of details including for example accounting, purchasing and payroll details. Sales and marketing database 596 keeps track of items such as sales and marketing details, client details, customer relations management and stock management. Internal data warehouse 560 receives information from databases 590, 596 and 597. This internal data warehouse 560 will only be accessed by authorized internal users conducting legitimate business activities. A secure (internal) data warehouse 545 services the needs of Class Three users 515. Specific (and confidential) information 580 is extracted from internal data warehouse 560 that is then stored in secure customer data warehouse 545 where authorized users 515 can query 535 (for example as interrogation requests), specific and confidential information such as clinical history information, pathology results and interpretations. This information is presented in a secure user-friendly and/or visual format 585 in relation to individuals or groups of athletes or performance animals, and/or time series of results.

Figure 13 is a flow diagram of one specific example showing steps for assessing a biological sample for diagnosing or assessing a condition of an animal. A user collects a

biological sample 1010, for example a blood sample from a horse. At the same time, biological parameters including biochemical and haematological parameters, clinical data (including blood profile tests) and appraisal information are collected and recorded in a standard format 1015, for example by filling in a standard form. The biological sample 1010 is processed so that nucleic acids contained therein are detectable when hybridised with a complementary (or mismatch-complementary) nucleic acid located on an array 1020. The nucleic acid may be detectable by a label incorporated therein, for example a target nucleic acid. Preferably, the array 1020 is a device such as a microarray which is read 1030 by standard methods and equipment common to the art to identify and measure relative abundance or absolute abundance of those nucleic acids from the biological sample which have bound to probe nucleic acids immobilised as part of array 1020 (inclusion of a reference sample run in parallel allows for the calculation of the relative abundance of target nucleic acids, whereas a method developed by the company Affymetrix, Inc (the "Affymetrix system") as described at their website "affymetrix.com" relies on internal references).

Array 1020 may comprise a large number of probe nucleic acids, eg. 1000's of nucleic acids. A large number of probe nucleic acids may be particularly useful if an animal is not presenting with any visible signs of poor condition, eg. overt disease. Accordingly, in one embodiment, labelled target nucleic acids of a sample are first applied to an array comprising a "full-screen" of target nucleic acids (eg. 1,000's of nucleic acid probes that represent most or many of the nucleic acids expressed in a sample). Based on results from the full-screening, the labelled nucleic acid targets may be applied to a sub-set of the full-screen, eg. a selected panel of nucleic acid targets that may be associated with a particular condition, for example, respiratory diseases, drug consumption, etc.

Data from the read microarray 1030 and clinical data and appraisal information 1015 is formatted 1040 and transmitted via a communications network 1050, for example the Internet, to a remote diagnostic server 1060. It will be appreciated that transmission of the formatted data to the remote diagnostic server 1060 requires less bandwidth than transmitting database information to the user and less skill and time on behalf of the user.

The transmitted data is analysed 1070, for example by comparison to a database of previously collected information in relation to clinical information and expression levels (relative abundance) of the nucleic acids applied to the microarray 1020. Also, experts, for example, bioinformaticists, biologists, doctors, pathologists, and the like may analyse the data to provide additional useful information. The analysis enables correlation to a condition 80. In this manner, the expression levels (relative or absolute abundance) of the nucleic acid probes applied to the microarray 1020 are correlated with previously collected data relating to known conditions stored in a database 1080 and compiled 1090. The database may also store information in relation to an identity of known nucleic acids, nucleotide sequence on the array and/or location of nucleic acids on the array, its biological function and links to other databases.

Results in relation to health and performance condition are transmitted via a communications network 1050 and may also be provided to the user as a report 1095, for example a hardcopy printout or visually on a computer monitor.

The described system has advantages of requiring low bandwidth for transmitting sample data and final report between user and remote database/processor, data processing is centralised and more efficient, expert analysis of the sample data is centralised, the computer software may incorporate heuristic methods thereby minimising human interaction, the possibility of user and interpretation bias is avoided, and information stored in the commercially valuable database is under strict control and does not require direct access by an outside user. The steps are described in more detail hereinafter.

Figure 14 shows an environment for working the method described in Figure 13. A user 1100, which may be a veterinarian or practitioner, collects a sample 1120 from an animal 1101, for example a blood sample from a horse or athlete. Concurrently, information in relation to a condition of the animal is collected in a standard format 1102. The sample is collected, nucleic acids isolated therefrom, prepared and applied to an array 1120 and the array is read by an array reader 1130. Data from the array reader 1130 and clinical appraisal and condition information 1102 is entered into a computer and formatted by a

processor 1140, which may be for example, a laptop computer with a modem. The formatted data is transmitted via a communications network 1150, for example the Internet. A remote diagnostic server 1160 receives the transmitted data and the data is compared with a database(s) 1161 which stores data, for example, data in relation to nucleic acid location on an array, expression level (relative abundance or absolute abundance) of a nucleic acid hybridised with a corresponding nucleic acid on an array, and data correlating nucleic acid expression level and performance, health, or condition of an animal.

Figure 15 is a flow diagram illustrating steps for preparing an array. A biological sample 1210 is collected from an animal. Biological sample 1210 may comprise for example, a blood sample (preferably white blood cells isolated therefrom), urine sample or tissue sample (including fetal tissues and tissues in various stages of development). A specific aim of collecting the biological sample is to isolate and sequence as many relevant genes from the sample for use on an array. Thousands of nucleic acids may be isolated that may form a large number of probes for a broad screening of an animal's genetic make-up or gene expression pattern.

Nucleic acids are isolated from the biological sample. In one instance the sample may be used to prepare genomic DNA or tissue specific mRNA 1223. In another instance RNA is isolated from the biological sample 1210 and a cDNA library 1220 is prepared from the isolated RNA. Plasmids 1221 comprising cDNA inserts from library 1220 may be sequenced 1222 from either or both 5' and/or 3' end of the nucleic acid. Preferably, sequencing is from the 3' end. Sequences may comprise Expressed Sequence Tags (EST). If an isolated nucleic acid does not encode a full-length gene (eg. an EST), a partial nucleic acid may be used as a probe to isolate a full-length nucleic acid. Alternatively, or in addition, EST sequence information may be compared directly with a sequence database 1230, for example GenBank, and a search for related or identical sequences performed. Putative gene identification and function 1231 may be determined from a search, for example a BLAST search performed in step 1230. By determining the number of times

each gene is represented in the library, a computer may be programmed to enable the normalisation and standardisation of the relative abundance data of mRNAs in a sample.

Gene-specific oligonucleotides 1232 may be synthesised using information from EST or full-nucleotide sequence 1222 data. Gene-specific oligonucleotides 1232 may be used as amplification primers to amplify (step 1224) a region of a corresponding nucleic acid. The nucleic acid used as template to amplify a region of corresponding nucleic acid may be, for example, isolated plasmid DNA 1221 and/or genomic DNA, cDNA or mRNA (eg. used with RT-PCR) 1223. The nucleic acid thus prepared can be used directly as the nucleic acids for attaching to an array 1240. Amplification products 1225 may also be generated using non-gene-specific primers (eg. oligo-dT, plasmid sequence flanking a nucleic acid of interest). Oligonucleotides corresponding to a gene 1232 may also be used on array 1240, alternatively the oligonucleotide corresponding to known sequence can be built successively nucleotide by nucleotide on a support using Affymetrix methodology such as that in US patent no. 5,831,070, incorporated herein by reference.

In one embodiment, the step relating to constructing cDNA 1220 and isolating plasmids 1221 comprising the cDNA may be omitted. In this embodiment, isolated genomic DNA or tissue specific mRNA 1223 is used as a template to make amplification product 1225 by amplification using gene-specific primers 1232. Amplification product 1225 may be attached to array 1240.

Nucleic acids attached to or built onto array 1240 preferably represent most, more preferably all, expressed genes in a given tissue from an animal of interest. For example, for a complete diagnostic test for racehorse blood, the array should contain genes expressed in the cells of blood under various conditions and at various stages of cell differentiation.

Figure 16 shows a flow diagram comprising steps for determining gene expression in biological samples comprising both reference target 1305 and sample target 1310. Nucleic acids, in particular RNA (total RNA or mRNA), are isolated from biological samples 1305

and 1310, which may be the same sample. cDNA is prepared from the RNA and the cDNA is labelled resulting in labelled targets 1320 and 1325. Alternatively, or in addition, cDNA may be used as a template to synthesise labelled antisense RNA for use as targets 1320 and 1325. Reference target 1325 may be provided as a previously prepared labelled target of known concentration. Accordingly, reference target 1325 need not be synthesised in parallel with each sample target. Internal controls for reference target 1325 and sample target 1320 provide a means for normalising and scaling relative probe concentrations.

Sample target 1320 and reference target 1325 are hybridised with array 1330 in step 1340. Array 1330 may, for example, have been prepared by steps shown in Figure 15. The hybridised array is washed 1345 to remove non-specific hybridisation of targets 1320 and 1325. It will be appreciated that one skilled in the art could select different stringency conditions of wash 1345 as required. Array 1330 is read in an array reader 1350 to determine relative abundance of RNA in the original sample, which correlates with expression of the corresponding gene in the biological sample.

Figure 17 is a flow diagram illustrating steps for building a database. Biological samples 1410 are collected from animals having specific known condition(s). Preferably, a statistically relevant number of biological samples 1410 are collected from a variety of normal animals to establish a normal reference range of nucleic acid abundance levels. This should account for natural variation, including that associated with state of fitness, sex, age, season, breed and diurnal changes. Nucleic acids are isolated and labelled 1415 from sample 1410, thereby forming respective target nucleic acids. The labelled target nucleic acids 1415 are applied to array 1420, which may be prepared as described in Figure 15. The array is read 1430 and data formatted 1440 into an electronic form, for example a digital signal, suitable for transmission via a communications network 1450. Clinical information from clinical appraisal, in relation to conditions of animals of interest is measured, documented and compiled 1460. The clinical information is preferably collected in a standard format, and for example, variable states such as the level of fitness or body score (fatness) may be assigned given a value or number (for example between 1-10). Specific clinical conditions may be graded (for example between 1-10) and assigned a

unique and standard identifier. An example of such a system is currently used in clinical medicine and veterinary science and termed SNOMED or SNOVET (Standardised Nomenclature of Medicine or Veterinary Science), where a clinical condition can be described using a numerical system. This system has not been used for describing the normal condition or the ability of a performance animal to perform to its best. A numerical grading system could also be used to standardise the collection of such data, for example, time spent on a treadmill is a strong indicator of exercise tolerance, as is blood concentration of oxygen and ability to transport oxygen. Conditions may include disease, response to drugs, training, nutrition and environment. The clinical information 1460 is formatted into electronic form 1440, for example a digital signal, suitable for transmission via a communications network 1450.

The process is repeated such that a collection of several array readouts for particular conditions are made. A standard range (for example, a population median of 95%) of values for each of the represented genes and its relative abundance can be calculated. This reference range can then be used as a comparison to test sample results.

Nucleic acid expression information from a read array 1430 for a target sample is correlated with previously measured conditions 1460 to provide information on nucleic acid expression level (abundance or relative abundance) with any previously measured condition. This information is compiled at server 1470 and good data is stored and bad data rejected 1480. The compilation process includes collection of a large enough set of array readout information for a particular condition so that inferences can be drawn on gene expression profiles and conditions. The compilation 1470 may also include use of sophisticated pattern recognition and organisational software and algorithms (examples common to the art include algorithms such as K means, Anova and Mann Whitney, Self Organising Maps, principal component analysis, hierarchical clustering – any one of which is available as part of proprietary software packages) such that expression patterns that differ to normal or expected condition can be identified. The compilation 1470 will preferably include sophisticated methods of supervised classification such as regularised discriminant analysis, diagonal discriminant analysis, support vector machines, or

recursive partitioning – any one of which is readily conducted using proprietary software packages. Concurrently, comprehensive clinical information 1460 for animals may be collected and biological samples 1410 tested on arrays so that correlations can be made between any clinical observation and array data. In this manner a database is created comprising data on nucleic acid expression which may include data correlating any desired condition, for example normal and specific abnormal condition(s), with nucleic acid expression. The stored data 1480 may be accessed using specific programs and algorithms 1490.

Throughout this specification, unless the context requires otherwise, the words comprise, comprises and comprising will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

In order that the techniques outlined above may be readily understood and put into practical effect, particular preferred embodiments will now be described by way of the following non-limiting examples.

STEP 1

Biological Sample Collection

A biological sample comprising nucleic acids, for example total RNA and mRNA, is collected. The biological sample may include cells of the immune system at various stages of development, differentiation and activity. The biological sample in most instances would be whole blood collected from a vein of a performance animal. However, the biological sample may include a fluid and/or tissue, for example sputum, urine, tissue biopsies, bronchial or nasal lavages, joint fluid, peritoneal fluid or thoracic fluid which, in part, comprises cells of the immune system that have infiltrated such tissues or fluids. Cells present in blood which comprise mRNA may include mature, immature and developing neutrophils, lymphocytes, monocytes, reticulocytes, basophils, eosinophils, macrophages. All of these cell types also appear in tissues of non-blood origin at various times in various conditions.

Methods described herein may include use of the abovementioned cell types. The biological sample is collected and prepared using various methods. For example, an easy method of collecting cells of the blood is by venipuncture. The biological sample may be collected from a performance animal, for example, a horse with suspected laminitis, a human athlete or camel with osteochondrosis, or a greyhound with subclinical cystitis.

Blood sample

Ten ml of blood is drawn slowly (to prevent hemolysis) from the vein of an animal (jugular vein in a horse and camel, veins on the forearm/limb of humans and dogs) into a 1:16 volume of 4% sodium citrate to prevent clotting and the sample is mixed and then placed on ice. The sample is centrifuged at 3000 RPM at 40C for 15 minutes and white blood cells (WBC) (commonly called the "buffy coat") are removed from the interface between plasma and red blood cells (RBC) into a separate tube using a pipette. The WBCs are then treated with at least 20 volumes of 0.8% ammonium chloride solution to lyse any contaminating RBC and re-centrifuged at 3000 RPM at 40C for 5 minutes. The pelletted WBCs are then washed in 0.9% sodium chloride, re-centrifuged, and kept on ice. The cell pellet is then used directly in RNA extraction.

Non-blood biological fluid sample

A fluid sample, for example, sputum, urine, bronchial or nasal lavages, joint fluid, peritoneal fluid or thoracic fluid, is centrifuged at 3000 RPM at 4 0C for 20 minutes to collect cells. Samples comprising large amounts of mucous are treated with a mucolytic agent such as dithiothretol prior to centrifugation. A cell pellet is then washed in 0.9% sodium chloride, re-centrifuged and the cell pellet is used directly in RNA extraction.

Tissue biopsy

A tissue biopsy is frozen in dry ice or liquid nitrogen and crushed to powder using a mortar and pestle. The frozen tissue is then used directly in RNA extraction.

STEP 2

RNA Isolation

Total RNA and/or mRNA is isolated from a biological sample. Use of isolated mRNA rather than total RNA may provide results with less background and improved signal.

RNA is commonly isolated by skilled persons in the art, and examples of some methods for isolating mRNA are described below.

Commercially available kits, for example, Qiagen RNA and Direct RNA extraction kits, and RNA extraction kits produced by Invitrogen (formerly Life Technologies) and Amersham Pharmacia Biotech herein incorporated by reference, may be used by following the manufacturer's instructions. Key elements of these mRNA extraction protocols include use of an appropriate amount of sample, protection of the sample from RNAse contamination, elution of the sample from a column at 70OC and quantitation and quality checking in an agarose 0.7% gel and using an OD 260/280 ratio. About 0.2 gm (wet weight) of pelleted white blood cells or tissue is required for each mRNA extraction which will yield about 1-2µg of mRNA. Disposable gloves should be worn throughout the procedure, with frequent changes. Both the column and solution used for elution should be at 70OC.

RNA quantification and assessment of RNA size and quality include standard gel electrophoresis methods of running a small quantity of an RNA sample on an agarose gel with known standards, staining the gel with for example ethidium bromide to detect the sample and standards and comparing relative intensities and size of standard RNA and sample RNAs, comparison of the intensities of the ribosomal RNA bands. Alternatively, or in addition, RNA concentration in a solution may be determined by measuring absorbance at 260/280 nm in a spectrophotometer relative to known standards and calculated using known formulas.

cDNA Synthesis and Labelling

RNA prepared as described above may be synthesised to cDNA and labelled resulting in a labelled probe using kits provided by suppliers such as Amersham Pharmacia Biotech, Invitrogen, Stratagene or NEN, herein incorporated by reference. For example, a typical

reaction may comprise: template RNA, an oligo-dT primer and/or gene-specific primers, reverse transcriptase enzyme, deoxyribonucleic triphosphates (dNTP), a suitable buffer, and a label incorporated into at least one of the dNTPs. Such a reaction when combined with a method of amplifying the resultant cDNA is referred to as RT-PCR (reverse transcriptase-polymerase chain reaction). A specific example is provided below, but it should be noted that other methods of incorporation of label into DNA can be used and that such methods are under constant review and improvement, for example some methods include the incorporation of amino-allyl dUTP and subsequent coupling of N-hydroxysuccinate activated dye to increase the specific labelling of the DNA.

To anneal primer(s) to template RNA, mix 2µg of mRNA or 50-100 µg total RNA from respective test sample (Cy3) and reference sample (Cy5) in separate tubes with 4µg of a regular or anchored oligo-dT primer or gene-specific primers in a total volume of 15 µl (using purified water to make up the volume). (Regular oligo dT is 5'-TTT TTT TTT TTT TTT TTT TTT, anchored oligo dT is 5'-TTT TTT TTT TTT TTT TTT TTV N-3'), (where V=A, C or G; and N=A, C, G or T). Heat mixture to 65oC for 10 min and cool on ice. Add 15.0 µl of reaction mixture to respective Cy3 and Cy5 reactions.

The reaction mixture comprises of the following: 6.0 ul of 5X first-strand buffer, 3.0 µl of 0.1M DTT, 0.6 ul of unlabeled dNTPs, 3.0 ul of Cy3 or Cy5 dUTP (1 mM, Amersham), 2.0 ul of Superscript II (Reverse transcriptase 200 U/µL, Life Technologies) made to 15 µl with pure water. Unlabelled dNTPs are sourced from a stock solution consisting of 25mM dATP, 25 mM dCTP, 25 mM dGTP, 10 mM dTTP. 5X first-strand buffer consists of 250 mM Tris-HCL (pH 8.3), 375mM KCl, 15mM MgCl2). The mixture is incubated at 42oC for 1 hr. Add an additional 1 µl of reverse transcriptase to each sample. Incubate for an additional 0.5-1 hrs. Degrade the RNA and stop the reaction by adding 15µl of 0.1N NaOH, 2mM EDTA and incubate at 65-70oC for 10 min. If starting with total RNA, degrade the RNA for 30 min instead of 10 min. Neutralize the reaction by adding 15µl of 0.1N HCl. Add 380µl of TE (10mM Tris, 1mM EDTA) to a Microcon YM-30 column (Millipore).

Next add 60µl of Cy5 probe and 60µl of Cy3 probe to the same microcon. Centrifuge the column for 7-8 min. at 14,000 x g. Remove flow-through and add 450 µl TE and centrifuge for 7-8 min. at 14,000 x g (washing step). Remove flow-through and add 450 µl 1X TE, 20 µg of species-specific Cot1 DNA (20ug/ul, Life Technologies for human – Cot1 DNA is genomic DNA that has been denatured and re-annealed such that the concentration of the DNA and the time of re-annealing multiplied equals 1. Methods for making Cot1 DNA are common in the art), 20µg polyA RNA (10 µg/ul, Sigma, #P9403) and 20 µg tRNA (10 µg/ul, Life Technologies, #15401-011). Centrifuge 7-10 min. at 14,000 x g. The probe needs to be concentrated such that with the addition of other solutions required for hybridisation the volume is not excessive, or is suitable for use with a desired slide and cover slip size. Invert the microcon into a clean tube and centrifuge briefly at 14,000 RPM to recover the probe.

A nucleic acid may be labelled with one or more labelling moieties for detection of hybridised labelled nucleic acid (ie. probe) and target nucleic acid complexes. Labelling moieties may include compositions that can be detected by spectroscopic, photochemical, biochemical, immunochemical, optical or chemical means. Labelling moieties may include radioisotopes, such as 32P, 33P or 35S, chemiluminescent compounds, labelled binding proteins, heavy metal atoms, spectroscopic markers, such as fluorescent markers and dyes, magnetic labels, linked enzymes, and the like. Preferred fluorescent markers include Cy3 and Cy5, for example available from Amersham Pharmacia Biotech (as decribed above).

cRNA synthesis and labelling

The Affymetrix system uses RNA as substrate and generates biotin labelled cRNA through a series of reactions detailed in a protocol available from their website (affymetrix.com), incorporated herein by reference. The cRNA is fragmented prior to application onto the array.

STEP 3

Arrays

One feature is an array comprising nucleic acids representing expressed genes from cells found in blood of a performance animal, for example a horse, human, camel or dog. The nucleic acids may be of any length, for example a polynucleotide or oligonucleotide as defined herein.

Each nucleic acid occupies a known location on an array. A nucleic acid target sample probe is hybridised with the array of nucleic acids and an amount or relative abundance of target nucleic acid hybridised to each probe in the array is determined.

High-density arrays are useful for monitoring gene expression and presence of allelic markers which may be associated with disease. Fabrication and use of high density arrays in monitoring gene expression have been previously described, for example in WO 97/10365, WO 92/10588 and US Patent No. 5,677,195, all incorporated herein by reference. In some embodiments, high-density oligonucleotide arrays are synthesised using methods such as the Very Large Scale Immobilised Polymer Synthesis (VLSIPS) described in US Patent No. 5,445,934, incorporated herein by reference.

Arrays for human are commercially available from companies such as Incyte, Research Genetics, and Affymetrix. Lion Bioscience recently announced forthcoming release of a dog microarray and have a clone collection of dog cDNAs. These arrays typically comprise between 2,000 and 10,000 genes and are species specific. None are available for the horse or camel. Some of these genes are in multiple copies on the array and have not been fully annotated or given a true gene identity. Additionally, it is not known whether DNA on the array, when hybridised to a test sample, specifically binds to a single gene. This latter instance results from splice variants of RNA transcripts in tissues such that one gene may encode multiple transcripts.

Human and dog arrays (when available) can be used in methods described herein. However, these arrays are currently non-specific and include genes that are not expressed in blood cells of animals, and/or do not contain genes important in controlling the function of blood cells, and/or contain regions of genes that are not specific to blood cells.

Clones containing specific genes are available and can be purchased for human (mouse and dog) for use on arrays (for example from the IMAGE consortium or Lion Bioscience). However, it is not possible to obtain specific clones for use on a blood-specific array without prior knowledge of what genes are expressed in blood cells. The IMAGE consortium also does not guarantee that the gene of interest is contained in the clone purchased.

Array Construction

Because of difficulties, problems and a likelihood of wasting financial resources to obtain a blood-specific DNA array, a method is provided herein which provides rapid and cost effective generation of species and tissue-specific DNA arrays for assessing nucleic acid expression in a sample. Figure 14 shows steps for constructing an array in one embodiment.

Target Nucleic Acid Preparation

Biological samples are collected as described above. Samples comprising cells expressing as many genes of interest in relation to condition(s) of a performance animal are collected. For example, a sample comprising a mixture of nucleated blood cells from performance animals with conditions such as, osteochondrosis, laminitis, tendon soreness, bursitis, abcesses, inflammation, allergy, viral infection, parasite infection, asthma, etc.

Approximately 5 μg of mRNA is isolated from the biological sample (typically 1 gm wet weight) using mRNA isolation kits or the protocol described above. Concurrently, 5 μg of mRNA is isolated from umbilical cord blood, and/or early stage foetus. Cells and tissues contained within these sources would express genes that may not be expressed in the cells extracted from blood in the above example. Isolation of cytoplasmic mRNA from cells is preferred. This step involves rupturing the cells with a solution comprising detergent and/or chaotropic agent and salt such that cell nuclei and the nuclear membrane remain intact. The cell nuclei are pelleted by centrifugation and the supernatant is used for mRNA extraction. Protocols for this procedure are available as part of mRNA isolation kits (eg

C:\WINNT\Profiles\VPA\Desktop\[illegible]

available by Qiagen). These mRNAs may be used to construct cDNA libraries. Kits for the construction of cDNA libraries are available from companies including Stratagene and Invitrogen (eg Uni-ZAP XR cDNA synthesis library construction kit #200450). The library preferably should be constructed such that the orientation of the cDNA in the vector is known, that the mRNA is primed using oligo dT, the vector is capable of receiving a nucleic acid insert up to 10 kb and that purification of DNA suitable for DNA sequencing is possible and easy. By following the manufacturer's instructions and paying particular attention to the quality of mRNA used and the size fractionation of cDNA (greater than 0.7 kb), a quality library containing enough viruses (>1x106) with insert sizes >0.7 kb can be generated.

Plasmids generated from such a library can be DNA sequenced using protocols that are well established in the art and are available, for example, from Applied Biosystems. Briefly, a mix of 0.5 μg of plasmid DNA, 3.2 pmol of a primer that hybridises to the vector DNA (eg M13 –21, or M13 reverse primer), thermostable DNA polymerase, dNTP and labelled dNTP is subjected to a routine PCR procedure to generate fragments of DNA that can be separated by gel electrophoresis and using machinery such as that available from Applied Biosystems (eg a 3700 DNA sequencer). Generated DNA sequence data (chromatogram) is assessed and quality scores and binning of similar sequences is done using a computer program package such as Phred/Phrap/Consed. The raw DNA sequence data can then be loaded into a database where comments (annotation) on the sequence can be made, such as quality score, bin, length of poly A sequence (should there be one), BLAST search results, highest homology in Genbank, clone identity, other entries in Genbank.

Subjective factors influencing whether a nucleic acid should be used on an array include quality and confidence of the DNA sequence, a Genbank homology score with identified nucleic acids, evidence of a poly-A tail (indicative of a translated transcript), uniqueness of the 3' sequence data (compared to both Genbank and an in-house database of clone sequences).

Nucleic acid primers can be selected using a program such as Primer 3 available via the Internet (www-genome.wi.mit.edu/cgi-bin/primer/primer3). The selected primers may be used for amplifying a nucleic acid, for example by PCR, or directly applied to an array. Uniqueness of a nucleic acid can be tested by performing additional BLAST searches on Genbank and an in-house database. Primers are preferably designed such that melting temperatures are similar, and amplification products are of a similar nucleic acid length. Primers for PCR are generally between 18 and 25 nucleotide bases long. Primers for direct use on a microarray or device are preferably between 50 and 80 nucleotide bases long. Both the amplification product and the single primer should hybridise to DNA that uniquely identifies a gene transcript. Specific programs using various formulas are available for calculating the melting temperature of various lengths of DNA (eg Primer 3). Alternatively, selected DNA sequences can be provided to Affymetrix for production of a proprietary and custom array. The sequences generated in-house are provided to Affymetrix in Fasta format along with details of which parts of the sequence to be used for the generation of a probe set (11 probes, each 25 nucleotide bases long) for each gene represented on the array.

Nucleotide sequences may be compared with an existing database, for example Genbank, to determine a previously provided name, tissue expression, timing of expression, biochemical pathway, cluster membership, and possible function or cellular role of an expressed nucleic acid. In addition, a nucleic acid fragment may be used as a probe to isolate a full-length nucleic acid which may encode a gene which is associated with a particular disease or condition. Further, identified nucleic acids may be used to isolate homologues thereof, inclusive of orthologues from other species. An identified nucleic acid may also be cloned into a suitable expression vector to produce an expressed polypeptide in vitro, which may be used, for example as an antigen in generating antibodies and for use on protein arrays. The antibodies may be used for developing specific diagnostic assays or therapies, for three-dimensional protein structure such as X-ray crystallographic studies, or for therapeutic development.

C:\WINNT\Profiles\VPA\Desktop\General\genetrix_as received_final.doc 14/11/02

An array may comprise any number of different nucleic acids, but typically comprises greater than about 100, preferably greater than about 1,000, more preferably greater than about 5,000 different nucleic acids. An array may comprise more than 1,000,000 different nucleic acids. Each nucleic acid is preferably represented more than once for scanning internal comparison and control. Preferably, the nucleic acids are provided in small quantities and are gene-specific and/or species-specific usually between 50 and 600 nucleotides long, arranged on a solid support.

The Affymetrix system uses 11 probes per gene, each of 25 nucleotides, that are built onto the array using a photolithographic method (US Patent Nos. 6,309,831; 6,168,948; 5,856,174; 5,599,695; 5,831,070; 6,153,743; 6,239,273; 6,271,957; 6,329,143; 6,310,189 and 6,346,413). The nucleic acids may be dotted onto the solid support or bound to microspheres, or in solution. A typical array may have a surface area of less than 1 cm2, for example a microarray.

A nucleic acid can be attached to a solid support via chemical bonding. Furthermore, the nucleic acid does not have to be directly bound to the solid support, but rather can be bound to the solid support through a linker group. The linker groups may be of sufficient length to provide exposure to the attached nucleic acid. Linker groups may include ethylene glycol oligomers, diamines, diacids and the like. Reactive groups on the solid support surface may react with one of the terminal portions of the linker to bind the linker to the solid support. Another terminal portion of the linker is then functionalised for binding the nucleic acid. A solid support may be any suitable rigid or semi-rigid support, including charged nylon or nitrocellulose, chemically treated glass slides available from companies such as NEN, Corning, S&S, arrays available through Affymetrix, membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The solid support can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which the nucleic acids are bound. Preferably, the solid support is optically transparent.

The array may be constructed using an "arraying machine" manufactured by companies for example Molecular Dynamics, Genetic Microsystems, Hitachi, Biorobotics, Amersham, Corning. Alternatively, the array may be manufactured according to specific instructions provided by the user to Affymetrix. Source materials for this machine include microtitre plates comprising nucleic acids representative of unique genes, or sequence information. An array element may comprise, for example, plasmid DNA comprising nucleic acids specific for a gene sequence, an amplified product using gene-specific or non-specific primers and template DNA or RNA, or a synthesised specific oligonucleotide or polynucleotide. Array elements may be purified, for example, using Sephacryl-400 (Amersham Pharmacia Biotech, Piscataway, N.J.), Qiagen PCR cleanup columns, or high performance liquid chromotography (for oligonucleotides).

Purified array elements may be applied to a coated glass substrate using a procedure described in U.S. Pat. No. 5,807,522, incorporated herein by reference. By other example, DNA for use on Corning amino-silane coated slides (CMT-GAPSTM) is re-suspended in 3xSSC to a concentration of 0.15-0.5 μg/μl and then used directly in an arraying machine in 96 or 384-well plates.

An example for preparing an array element is provided by the manganese superoxide dismutase gene. A clone comprising a nucleic acid insert is prepared and isolated as described above. The clone is sequenced to identify the nucleotide sequence. A BLAST search using the identified nucleotide sequence is performed to determine homology of the cloned nucleic acid with nucleic acids in a database, for example GenBank. Identification of nucleotide sequence homology with superoxide dismutase genes stored in the database provides a level of confidence that the clone comprises at least in part a gene for superoxide dismutase for the horse. Unique primers can be designed to amplify a nucleic acid using PCR and the clone DNA, or genomic DNA from the same species as a template. Purified amplification product can be directly attached to an array and thereby act as a target for a complementary labelled nucleic acid probe in the test and reference samples. Alternatively, a unique sequence can be determined and an oliognucleotide manufactured

C:\WINNT\Profiles\VPA\Desktop\Documents\patents_ [illegible]

and purified for direct use on an array, or the sequence information supplied directly to Affymetrix for the construction of a custom array.

The array may comprise negative and positive control samples (preferably as duplicates or triplicates) such as nucleic acids from species different from a sample being tested (negative controls) and various nucleic acids (representative of RNAs and both ends of RNA molecules) that are found in all tissues as a constant and known quantity (positive controls). These controls are identified and used by the array reader to provide data on true signal (ie. Specific hybridisation between probe and target) and noise (ie. Non-specific hybridisation between probe and target) and average intensity from multiple reads of several different locations for each nucleic acid attached to the array.

A test sample and a reference sample may be simultaneously assayed on the array. The reference sample may comprise mRNA from multiple sources, such that most, preferably all of the nucleic acids on the array are represented in the test sample, and can be used by the array reader as a non-zero standard and for comparison with an average of the read-outs from the test sample. A relative intensity for each gene on the array can be calculated.

The relative abundance of expression of each gene in a sample can also be calculated using controls within the array, such as certain genes expressed in a tissue at a constant level under all conditions.

Alternatively, using the Affymetrix system, an absolute level of expression is calculated based on the difference between the perfect match and mismatch hybridisation for each of the 11 probes for each gene. Using such a process a gene is scored as present or absent and an absolute measure of intensity is given along with a p value.

The interpreted array may highlight only a few genes that are substantially different in expression between a test and reference sample. Alternatively, the overall pattern of expression may provide a "fingerprint" to characterise the way in which the original cells have responded to a particular condition of a performance animal. For example, the gene

for superoxide dismutase may be the only gene up-regulated in a particular condition, especially in conditions of inflammation, or a large number of genes may be up- and down- regulated in various conditions. It is this fingerprint, rather than specific knowledge of gene sequence or function that can be used as a marker for various conditions. It would be expected that fingerprints be useful across species barriers to include performance animals such as humans, horse, dog and camel.

The arrangement of nucleic acids on the array may be periodically changed and these arrays are then assigned a particular batch code that corresponds to a specific array comprising a specific nucleic acid arrangement. The ability to change the arrangement of nucleic acids on the array and knowledge of the exact arrangement may prevent other people from generating a database using the arrays described above. Using a batch code also enables tracking of manufacturers of the arrays in regards to the number of arrays produced. The batch code further enables validation of a user of the communication network or "internet" diagnostic method and system. Batch code can also identify a particular type of array used, should more disease-specific arrays be designed and manufactured.

An example of how an array may be prepared and analysed is described in Eisen and Brown (Methods in Enzymology, 1999, 303 179) and in US Patent No. 6,114,114, herein incorporated by reference. Chapter 22 of Ausubel et al. supra also describes methods and apparatus for use with arrays and is herein incorporated by reference.

Control samples may be respectively labelled in parallel with a test and reference sample. Quantitation controls within a sample may be used to assure that amplification and labelling procedures do not change a true distribution of nucleic acid probes in a sample. For this purpose, a sample may include or be "spiked" with a known amount of a control nucleic acid which specifically hybridises with a control target nucleic acid. After hybridisation and processing, a hybridisation signal obtained should reflect accurately amounts of control nucleic acid added to the sample. For such purposes, a microarray may have internal controls, for example a nucleic acid encoding a common gene expressed by

the performance animal with known expression levels and a nucleic acid encoding a gene from another species that is known not to hybridise to the test or reference sample. To improve sensitivity and specificity of the assay, blocking agents such as Cot DNA from the tested species may also be used.

STEP 4

Hybridising Sample Nucleic Acid Probes with an Array

Nucleic acid probes may be prepared as described above from a biological sample from a performance animal that has been assessed concurrently by physical inspection and/or blood tests or other method. Nucleic acid targets from a statistically relevant number of normal animals previously hybridised to arrays, and a reference range for each of the genes on the array is calculated and used as a normal reference range (for example a 95% population median). Results from a test sample from a test animal can be compared with the same genes as the normal reference to determine if the test sample falls within the normal reference range. Further, nucleic acid targets may also be prepared from biological samples from apparently normal animals, animals with overt disease, various progressive stages of disease, hitherto undiagnosed or unclassified conditions or stages of such conditions, animals treated with known amounts of drugs (legal or otherwise), animals suspected of being treated with drugs (legal or otherwise), animals under specific exercise regimes for the sake of performance, animals subjected to (intentional or not) various nutritional states and/or environmental conditions. Databases of information from the use of such samples and arrays are created such that test samples can be compared. The database will then contain specific patterns of gene expression for particular conditions.

Prior to hybridisation, a nucleic acid probe may be fragmented. Fragmentation may improve hybridisation by minimising secondary structure and/or cross-hybridisation with another nucleic acid probe in a sample or a nucleic acid comprising non-complementary sequence. Fragmentation can be performed by mechanical or chemical means common in the art.

A labelled nucleic acid target may hybridise with a complementary nucleic acid probe located on an array. Incubation conditions may be adjusted, for example incubation time, temperature and ionic strength of buffer, so that hybridisation occurs with precise complementary matches (high stringency conditions) or with various degrees of less complementarity (low or medium stringency conditions). High stringency conditions may be used to reduce background or non-specific binding. Specific hybridisation solutions and hybridisation apparatus are available commercially by, for example, Stratagene, Clontech, Geneworks.

Affymetrix have detailed a standard procedure for the hybridisation of probes with an array (as describe at their website, affymetrix.com, incorporated herein by reference), however, a typical method entails the following:

Adjust probe volume (prepared as above) to a value indicated in the "Probe & TE" column below according to the size of the cover slip to be used and then add the appropriate volume of 20XSSC and 10% SDS.

Cover Slip Size (mm)	Total Hyb Volume (l)	Probe & TE (l)	20x SSC (l)	10% SDS (l)
22 x 22	15	12	2.55	0.45
22 x 40	25	20	4.25	0.75
22 x 60	35	28	5.95	1.05

20xSSC is 3.0 M NaCl, 300 mM NaCitrate (pH 7.0).

Denature the probe by heating it for 2 min at 100oC, and centrifuge at 14,000 RPM for 15-20 min. Place the entire probe volume on the array under the appropriately sized glass cover slip. Hybridize at 65oC (temperatures may vary when using different hybridisation solutions) for 14 to 18 hours in a custom slide chamber (for example a Corning CMT hybridisation chamber #2551).

C:\WINNT\ProGen\VP\A\Desktop\Documents\generic_ps revision_final.doc-14/11/02

Washing the Array

After hybridisation, the array is washed to remove non-specific probe and dye hybridisation. Wash solutions generally comprise salt and detergent in water and are commercially available. The wash solutions are applied to the array at a predetermined temperature and can be performed in a commercially available apparatus. Stringency conditions of the wash solution may vary, for example from low to high stringency as herein described. Washing at higher stringency may reduce background or non-specific hybridisation. It is understood that standardisation of this step is required to produce maximum signal to noise ratio by varying the concentration of salt used, whether detergent is present (SDS), the temperature of the wash solution and the time spent in the wash solution.

A typical wash protocol consists of removing the slide from a slide chamber, removing the cover slip and placing the slide into 0.1%SSC (recipe provided above) and 0.1% SDS at room temperature for 5 minutes. Transfer the slide to 0.1% SSC for 5 minutes and repeat. Dry the slide using centrifugation or a stream of air. Equipment is available to enable the handling of more than one slide at a time (for example, slide racks).

STEP 5

Reading the Array

After removal of non-hybridised probe, a scanner or "array reader" is used to determine the levels and patterns of fluorescence from hybridised probes. The scanned images are examined to determine degree of hybridisation and the relative abundance of each nucleic acid on the array. A test sample signal corresponds with relative abundance of an RNA transcript, or gene expression, in a biological sample. Alternatively, an Affymetrix array is read and computer algorithms calculate the difference between hybridisation on perfect match and mismatch probes for each of the 11 probes sets for each gene. It then calculates a presence or absence, an absolute value for each gene and a p value for the absolute call.

Array readers are available commercially from companies such as Axon and Molecular Dynamics and Affymetrix. These machines typically use lasers, and may use lasers at

different frequencies to scan the array and to differentiate, for example, between a test sample (labelled with one dye) and the control or reference sample (labelled with a different dye). For example, an array reader may generate spectral lines at 532 nm for excitation of Cy3, and 635 nm for excitation of Cy5.

A relative quantity of RNA may be calculated by the array reader and computer for respective nucleic acids on the array for respective samples based on an amount of dye detected, average of duplicate samples for respective genes and subtraction of background noise using controls. The reader is pre-programmed to perform such calculations (using proprietary software supplied with the array reader, such as MAS 5.0 for the Affymetrix system and Genepix for the Axon Instruments reader) and with information on the location of each nucleic acid on the array such that each nucleic acid is given a readout value. Controls or reference samples providing a readout for particular nucleic acids that falls within standard ranges ensures correct integrity of the array and hybridisation procedures. Programs typically generate digital data and format it for transmission

STEP 6

Querying and Transfer of Digital Data to a Central Database

Generated data is transmitted via a communications network to a remote central database. A user having access to the gene expression data enters information in relation to a test sample into a standard diagnostic form such that it can be digitalised. The information will include clinical appraisal and blood profile results. The format of such information is standard globally such that details on clinical conditions may be based on numerical input and each field of entry can be digitalised. For example, body temperature field could be number 0001, a recorded temperature within normal range would receive the number 0, 0.5OC above what is considered to be the normal range for that species would receive a number 5, 1OC above normal range would receive 10. Some examples of conditions that may be scored or rated in such a fashion are provided below.

a) Body temperature.

b) Integument: eyes, sores, abcesses, wounds, insects/parasites, allergy, infection.

c) Cardio/Respiratory: eyes, nasal discharge, rales, viral/bacterial infection, allergy, chronic obstructive pulmonary disease, cough/wheeze, crepitous sounds in the thorax, epistaxis, auscultation sounds, heart sounds, capillary refill, mucous membrane colour.

d) Gastrointestinal: diarrhoea, colic/stasis, parasites, appetite level, drenching time and dose.

e) Reproductive: stage of pregnancy, abortion, inflammation, discharges.

f) Musculoskeletal: lameness, laminitis, bone or shin soreness, muscle soreness or tying up, tendon or ligament affected, level of pain, X-ray data, scintigraphy data, CAT scan data, bursitis, bruising, cramping or "tying up".

g) Blood test results: biochemistry, immunology, serology (viral, bacteriological, hormone levels), cell counts, cell morphology, pathologist interpretation.

h) Other diagnostic test results: X-ray, biopsy, histopathology, CAT scan, MRI, bacteriology, virology.

i) Other data: Season (date), location, male or female, vaccination history, body score (fitness and fat), fitness level.

Alternatively, the entire system could be based on the aforementioned SNOMED system with appropriate modifications to encompass descriptions of exercise physiology and the normal animal. Alternatively, the entire system could rely on text or categorical data that can be appraised and scored by software such as Omniviz. Whatever system is used, if would be appreciated that the aim is to adequately, systematically and in a standard manner describe the current condition of the animal to the best of currently available technologies and could include results from machinery such as X-ray, ultrasound, scintigraphy and blood analysis.

The user also ensures that array results (that may for example be automatically collected from a reader), array specifications, data mining specifications, level of interpretation required and the clinical information are entered and correspond to the same animal and the same sample. The form is transmitted electronically to a central database and recognised as an individual accession or request by the database. The central database

recognises the user (using for example digital certificates), the user recognises the central database, the array batch code and gene array order are verified, and the user is allowed access (which may be automatic) and automatic processing of the request is performed if security and billing information are adequate. The processing involves specific mining of central data and specific user requested information is retrieved and resent automatically.

The above steps may be automated so that a user need not be present to perform the tasks. In an automated specific example, gene expression data from an array reader may be transmitted via a communications network directly to a server which is connected to a central database. Additional information could be input by the user at a processor which is also linked to the array reader.

Automated Data Mining Using Sent Data (Heuristic Methods)

A central database interprets the array specifications (eg. nucleic acid order on a microarray), decodes the information transmitted, determines nucleic acid expression level in a biological sample and compares the expression level and patterns of expression with known standards or reference range. Various levels of database interpretation may be applied to the data transmitted, depending on the user requirements. Clusters of genes may be up-regulated or down-regulated in certain conditions and the database makes automated correlations to specific conditions by accessing various levels of database information.

Mining software such as Metamine (Silicon Genetics), ArraySCOUT (Lion Bioscience) can be used in this instance, and more advanced data mining technologies could be used to identify patterns and nearest neighbour information in data (such as products from AnVil Informatics Inc and OmniViz Inc). Further, software capable of taking rule-based instructions (such as that described by Pacific Knowledge Systems Sydney Australia in their "ripple down" technology) and having the ability to self learn (heuristics and neural network systems) such as that described in Khan et al. Nature Medicine 7 (6) 673, incorporated herein by reference, could be used at this stage to limit the level of human interaction in determining a diagnosis. In this latter example, an artificial neural network is used, and samples are divided into training and validation sets to create trained

calibrated models. The calibrated models are then used to rank genes in diagnostic importance.

Levels of database may include:

- Unique gene sequences (eg 3' and 5' EST sequence of genes)
- Gene identity, homologous genes, tissue expression, keywords, function, cellular role, gene clusters, biochemical pathway, PubMed references
- Primer sequences used to generate amplification products (eg two primer sequences used to uniquely amplify the gene for gamma interferon in a particular species)
- Microarray construction and format (eg coded information on array manufacture batch and identification of genes and position on the array)
- Blood profile and clinical data associated with particular conditions (eg standard clinical information and IDEXX-machine generated blood profile data)
- Array data for normal and apparently normal status (eg 95% median range for normal animals)
- Array data for inducible disease and disease models
- Array data for various overt diseases (eg joint inflammation)
- Array data for stages of various overt diseases (eg pre-clinical, clinical and recovery stages)
- Array data for the influence of various classes of drugs, legal or otherwise, of known administration and dose, or unknown administration or dose (eg various steroids)
- Array data for the response to known and various levels of drugs used as a therapy (eg various anti-inflammatory medication at specific doses for a specific condition)
- Array data for the response to exercise and various training regimes
- Array data for the response to nutrition and various feeding regimes
- Array data for the response to the environment so as to possibly determine influence of during various seasons, or allergens or feed types.

Each successive level relies on at least one previous level of database to allow for interpretation. The database may be built over time and more intensive searching of the

database may incur a greater cost. As the database grows, changes may be made to the above methodology to increase the sensitivity of the detection of variation in expression of condition-specific genes – this could include the use of condition-specific arrays or condition-specific primers. Condition-specific arrays can be manufactured by a company such as Affymetrix (under instructions) that would allow for increased sensitivity and specificity, much reduced size of arrays, decreased cost of production, and the ability to process multiple samples at once. The process of building the database is iterative, such that specific genes are correlated to specific conditions, and the detection of variations in these genes becomes more sensitive and specific through the use of various modifying processes through the procedure (eg. the use of gene-specific primers for the amplification and labelling of cDNA from RNA, and the selection of limited numbers of genes on a disease- or condition-specific array, detection of splice variants and single nucleotide polymorphisms).

STEP 7

Standardised Electronic Reporting

The database reports back electronically to a remote user, either automatically or with a level of human intervention. The electronic report may be converted to a printed document. The report provides details of an animal's condition that is determined by correlation of gene expression data with information stored in a remote database, and optionally expert analysis.

Information sent might include:

- Individual genes up-regulated or down-regulated (for example, with laminitis or joint capsule inflammation or bursitis, a report on the up-regulation of genes such as interleukin-3, manganese superoxide dismutase, Groα, metalloproteinase matix-metallo-elastase, ferritin light chain may have some correlation to tissue inflammation, and down-regulation of genes such as insulin-like growth factor and its receptor may be correlated to recovery from such a condition). The identity of these genes cannot be predicted to be associated to any condition unless the above described methodology is used and databases on relative expression of genes for particular conditions have been

compiled. Therefore a screening test covering all genes may need to be performed first and a second, more specific test then applied.

- The overall pattern of gene expression and any correlation to particular conditions. For example, animals in heavy training may have a gene "fingerprint" that is different to animals being spelled from training.
- Individual pattern of gene expression (ie. the shape of the gene expression pattern over a time course or multiple samples taken over a period may change as an animal recovers from a condition)
- Changes to a pattern of gene expression, gene expression profile or level for a single animal over a time period or for successive tests.
- Clusters of genes up-regulated or down-regulated in a particular condition
- Pathways of genes up-regulated or down-regulated in a particular condition
- Correlations between genes up-regulated or down-regulated and known conditions, or stage of condition, or influence
- Known therapies to ameliorate the condition or enhance desired effects
- Specialist pathologist written interpretation
- Relevant information of use to veterinarians, medical practitioners, owners, trainers and athletes
- Collections of data on groups of animals under specific management regimes

Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. It would therefore be appreciated by those of skill in the art that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present invention. For example, the examples described herein may be used with performance animals other than horse, for example human, dog and camel.

All references, inclusive of patents, patent applications, scientific documents and computer programs, referred to in this specification are herein incorporated by reference in its entirety.

Dated this of Fourteenth day of November, 2002

GENOMICS RESEARCH PARTNERS PTY LIMITED

By their Patent Attorneys

DAVIES COLLISON CAVE

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1) A method of determining the status of a subject, the method including:
   a) Obtaining subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;
   b) Comparing the subject data to predetermined data, the predetermined data including for each of a number of individuals:
      i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to distinguish each of the conditions; and,
      ii) An indication of the status of the respective individual; and,
   c) Determining the status of the subject in accordance with the results of the comparison.
2) A method according to claim 1, the number of parameters being greater than 100.
3) A method according to claim 1, the number of parameters being greater than 1000.
4) A method according to claim 1, the number of parameters being less than 6000.
5) A method according to claim 1, the method including determining any conditions displayed by the user.
6) A method according to claim 5, the method including determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.
7) A method according to claim 1, the method of performing the comparison including causing the second processing system to:
   a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,
   b) Use the templates to classify the subject data into a respective one of the groups.
8) A method according to claim 7, the method including determining one or more conditions in accordance with the determined groups.
9) A method according to claim 1, the parameters being representative of the level or abundance of a molecule selected from one or more of :

C:\WINNT\Profiles\V7A\Desktop\Documents\Applications\...

a) A polynucleotide;

b) A polypeptide; and,

c) A polysaccharide.

10) A method according to claim 1, the method including:

a) Receiving confirmation of the determined status; and,

b) Updating the predetermined data in accordance with the confirmed status and the subject data.

11) A method according to claim 1, the predetermined data including phenotypic information of the individuals, and the subject data including phenotypic information regarding the subject, the phenotypic information including details of one or more phenotypic traits.

12) A method according to claim 11, the method including comparing the subject data to predetermined data for individuals having one or more phenotypic traits in common with the subject.

13) Apparatus for determining the status of a subject, the apparatus including a processing system adapted to:

a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;

b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to distinguish each of the conditions; and,

ii) An indication of the status of the respective individual; and,

c) Determine the status of the subject in accordance with the results of the comparison.

14) A computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a

suitable processing system causes the processing system to perform the method of claim 1.

15) A method of allowing a user to determine the status of a subject, the method including:

a) Receiving subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Causing the base station to:

i) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

(1) One or more parameter values for the respective individual; and,

(2) An indication of the status of each individual; and,

ii) Determine the status of the subject in accordance with the results of the comparison; and,

c) Transferring an indication of the status of the subject to the user via the communications network.

16) A method according to claim 15, the method including:

a) Having the user determine the subject data using a remote end station; and,

b) Transferring the subject data from the end station to the base station via the communications network.

17) A method according to claim 15, the base station including first and second processing systems, the method including:

a) Transferring the subject data to the first processing system;

b) Transferring the subject data to the second processing system; and,

c) Causing the second processing system to perform the comparison.

18) A method according to claim 17, the method including:

a) Transferring the results of the comparison to the first processing system; and,

b) Causing the first processing system to determine the status of the subject.

19) A method according to claim 17, the method including at least one of:

a) Transferring the subject data between the communications network and the first processing system through a first firewall; and,

b) Transferring the subject data between the first and the second processing systems through a second firewall.

20) A method according to claim 17, the second processing system being coupled to a database adapted to store the predetermined data, the method including:

a) Querying the database to obtain at least selected predetermined data from the database; and,

b) Compare the selected predetermined data to the subject data.

21) A method according to claim 17, the second processing system being coupled to a subject database, the method including storing the subject data in the subject database.

22) A method according to claim 15, the status including details of any conditions of the individuals, the method including determining any conditions displayed by the user.

23) A method according to claim 22, the method including determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.

24) A method according to claim 22, the method of performing the comparison including causing the second processing system to:

a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,

b) Use the templates to classify the subject data into a respective one of the groups.

25) A method according to claim 24, the method including determining one or more conditions in accordance with the determined groups.

26) A method according to claim 15, the parameters being representative of the level or abundance of a molecule selected from one or more of:

a) A polynucleotide;

b) A polypeptide; and,

c) A polysaccharide.

27) A method according to claim 15, the method including having the user determine the subject data using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number, the method including causing the base station to:

a) Determine the serial number from the subject data;

b) Determine a layout indicating the position of each feature on the array;

c) Determining the parameter values in accordance with the determined layout, and the subject data.

28) A method according to claim 15, the method including:

a) Receiving confirmation of the determined status from the user; and,

b) Updating the predetermined data in accordance with the confirmed status and the subject data.

29) A method according to claim 27, the features including at least one of:

a) An oligonucleotide;

b) A peptide; and,

c) An antibody.

30) A method according to claim 15, the predetermined data including phenotypic information of the individuals, and the subject data including phenotypic information regarding the subject, the phenotypic information including details of one or more phenotypic traits.

31) A method according to claim 17, the method including comparing the subject data to predetermined data for individuals having one or more phenotypic traits in common with the subject.

32) A method according to claim 15, the method including causing the base station to:

a) Determine payment information, the payment information representing the provision of payment by the user; and,

b) Perform the comparison in response to the determination of the payment information.

33) A base station for determining the status of a subject, the base station including:

a) A store method for storing predetermined data, the predetermined data including for each of a number of individuals:

i) One or more parameter values, at least one of the parameters being indicative of the current biological status of the individual;

ii) An indication of the status of the individual; and,

b) A processing system, the processing system being adapted to:

i) Receive subject data from the user via a communications network, the subject data including one or more parameter values;

ii) Compare the subject data to the predetermined data;

iii) Determine the status of the subject in accordance with the results of the comparison; and,

iv) Output an indication of the status of the subject to the user via the communications network.

34) A base station according to claim 33, the processing system being adapted to receive subject data from a remote end station adapted to determine the subject data.

35) A base station according to claim 33, the processing system including:

a) A first processing system adapted to:

i) Receive the subject data; and

ii) Determine the status of the subject in accordance with the results of the comparison; and,

b) A second processing system adapted to:

i) Receive the subject data from the processing system; and,

ii) Perform the comparison; and,

iii) Transfer the results to the first processing system.

36) A base station according to claim 35, the base station including:

a) A first firewall for coupling the first processing system to the communications network; and,

b) A second firewall for coupling the first and the second processing systems.

37) A base station according to claim 35, the processing system being coupled a subject database, the processing system adapted to store the subject data in the subject database.

38) A method according to claim 35, the method of performing the comparison including causing the second processing system to:

a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,

b) Use the templates to classify the subject data into a respective one of the groups.

39) A method according to claim 38, the method including determining one or more conditions in accordance with the determined groups.

40) A base station according to claim 33, the subject data being determined using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number, the processing system being adapted to:

a) Determine the serial number from the subject data;

b) Determine a layout indicating the position of each feature on the array;

c) Determining the parameter values in accordance with the determined layout, and the subject data.

41) A base station according to claim 33, the processing system being adapted to:

a) Receive confirmation of the determined ability; and,

b) Update the predetermined data in accordance with the determined ability and the subject data.

42) A base station according to claim 33, the predetermined data including phenotypic information of the individuals, and the subject data including phenotypic information regarding the subject, the phenotypic information including details of one or more phenotypic traits.

43) A computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of claim 16.

44) An end station adapted to determine the status of a subject, the end station including a processor adapted to:

a) Determine subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Transfer the subject matter to a base station via a communications network, the base station being adapted to:

i) Compare the subject data to predetermined data for one or more individuals, the predetermined data including:

(1) One or more parameter values for the respective individual; and,

(2) An indication of the status of each individual; and,

ii) Determine the status of the subject in accordance with the results of the comparison; and,

c) Receive an indication of the status of the subject to the user via the communications network.

45) A computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to operate as an end station according to claim 44.

46) A method of determining the ability of a subject to perform in a sporting and/or racing event, the method including:

a) Obtaining subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Comparing the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more parameter values for the respective individual; and,

ii) An indication of the status of each individual;

c) Determining the status of the subject in accordance with the results of the comparison; and,

d) Providing an indication of the ability in accordance with the results of the comparison.

47) A method according to claim 46, the status of each individual indicating any conditions displayed by the user, the method including:

a) Determining any conditions displayed by the user in accordance with the results of the comparison; and,

b) Determining the ability in accordance with the determined conditions.

48) Apparatus for determining the ability of a subject to perform in a sporting and/or racing event, the apparatus including a processing system adapted to:

a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

i) One or more parameter values for the respective individual; and,

ii) An indication of the status of each individual;

c) Determine the status of the subject in accordance with the results of the comparison; and,

d) Provide an indication of the ability in accordance with the results of the comparison.

49) A computer program product for determining the ability of a subject to perform in a sporting and/or racing event, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of claim 46.

50) A method of providing secure arrays for use, each array including a number of predetermined features, the method including:

a) Determining a number of respective feature layouts, each layout representing the positioning of each feature on a respective array;

b) Determining a number of serial numbers, each serial number corresponding to a respective layout;

c) Generating a number of arrays, each array being generated in accordance with a respective layout, and including the corresponding serial number thereon, the serial number being used in processing used the array.

51) A method according to claim 50, the method being performed to provide the arrays on behalf of an entity, the method including providing an indication of the layouts and corresponding serial numbers to the entity, to thereby allow the entity to process the arrays.

52) A method according to claim 50, the method of determining the layouts including:

a) Determining a preferred layout; and,

b) Moving the position of one or more of the features from the position in the preferred layout to alternative position.

53) A method according to claim 52, the method including:

a) Determining the type of each feature; and,

b) Exchanging the position of one or more features having different feature types.

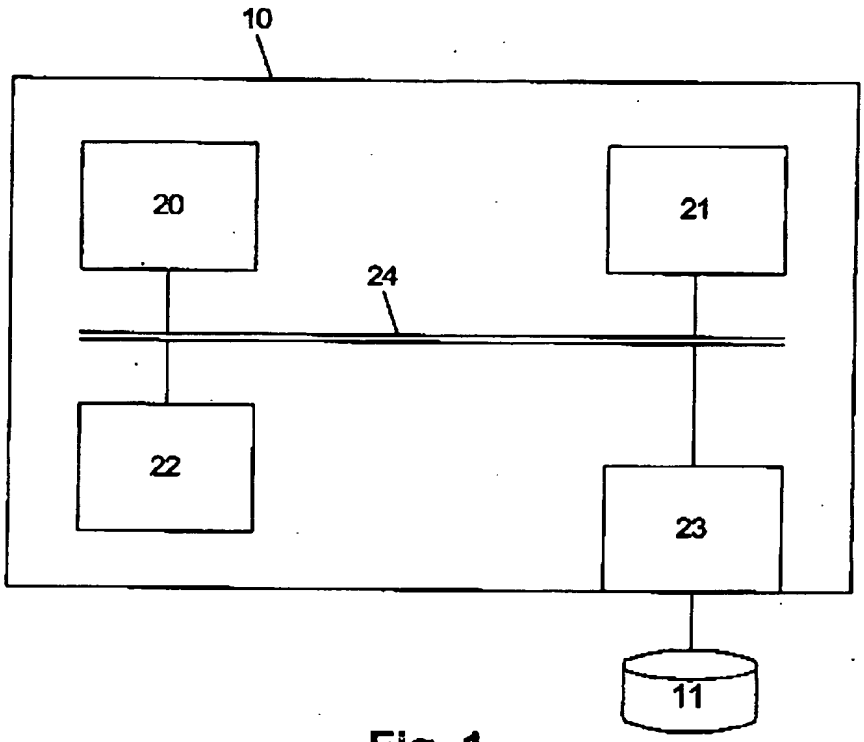

Fig. 1

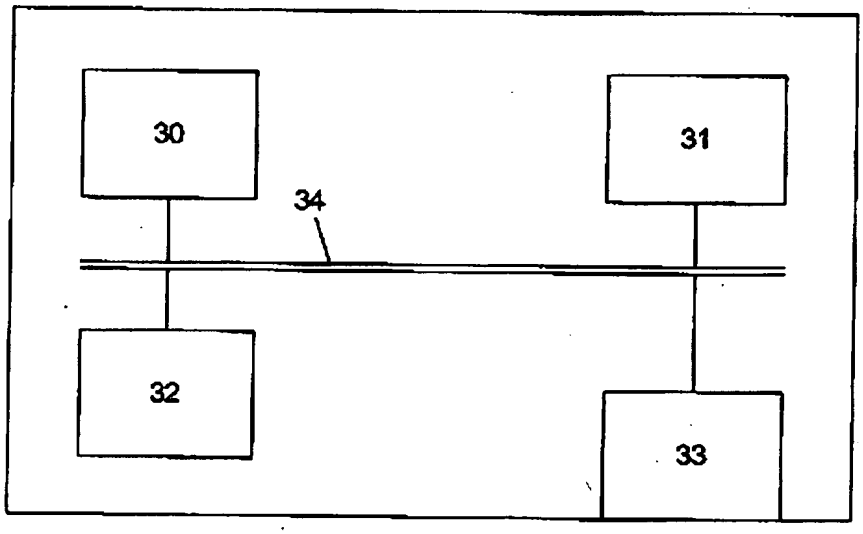

Fig. 4

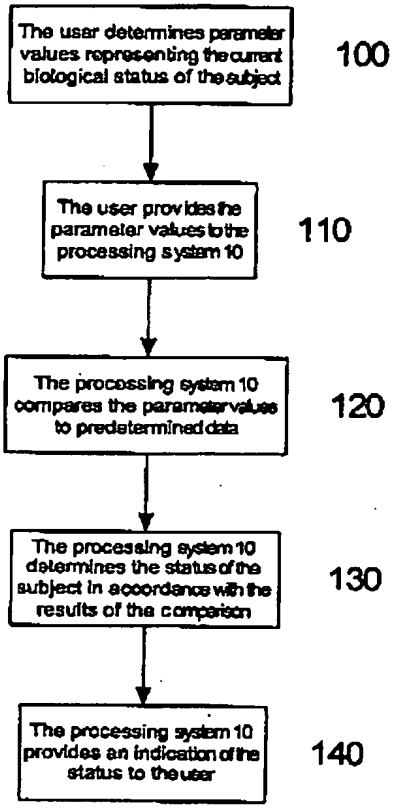
The user determines parameter values representing the current biological status of the subject
100
The user provides the parameter values to the processing system 10
110
The processing system 10 compares the parameter values to predetermined data
120
The processing system 10 determines the status of the subject in accordance with the results of the comparison
130
The processing system 10 provides an indication of the status to the user
140

Fig. 2

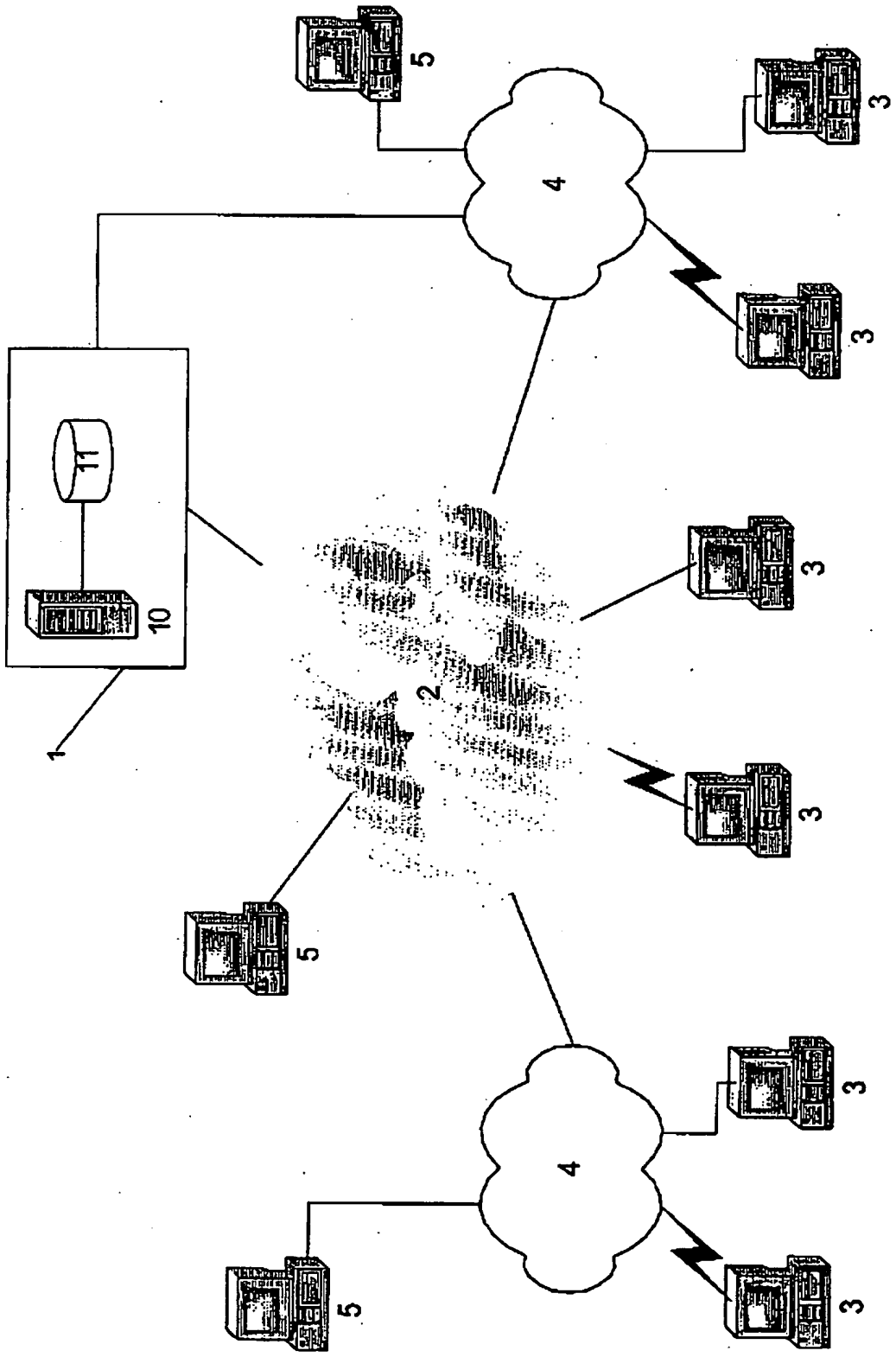

Fig. 3

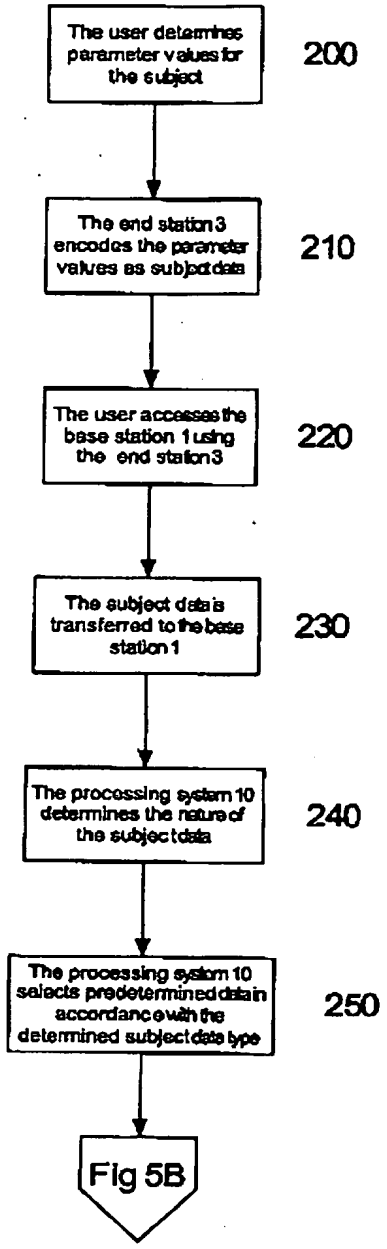
The user determines parameter values for the subject
200
The end station 3 encodes the parameter values as subject data
210
The user accesses the base station 1 using the end station 3
220
The subject data is transferred to the base station 1
230
The processing system 10 determines the nature of the subject data
240
The processing system 10 selects predetermined data in accordance with the determined subject data type
250
Fig 5B

Fig. 5A

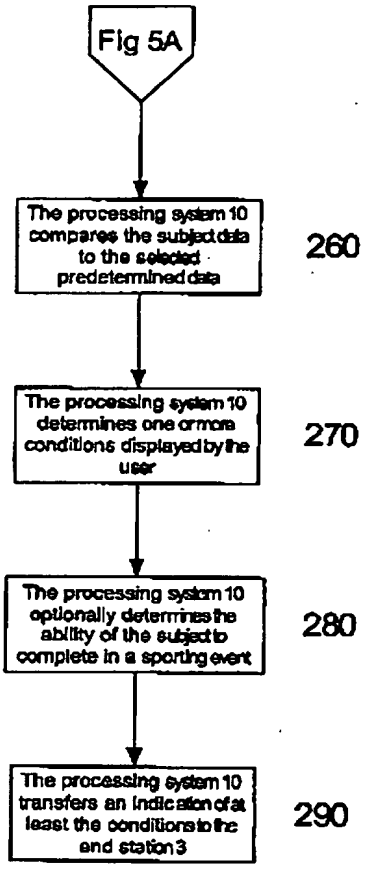
Fig 5A
The processing system 10 compares the subject data to the selected predetermined data
260
The processing system 10 determines one or more conditions displayed by the user
270
The processing system 10 optionally determines the ability of the subject to complete in a sporting event
280
The processing system 10 transfers an indication of at least the conditions to the end station 3
290

Fig. 5B

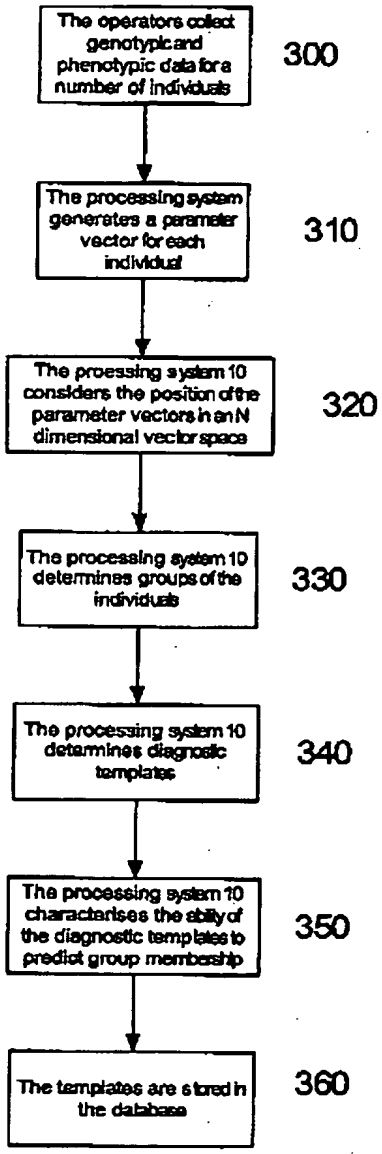
The operators collect genotypic and phenotypic data for a number of individuals
300
The processing system generates a parameter vector for each individual
310
The processing system 10 considers the position of the parameter vectors in an N dimensional vector space
320
The processing system 10 determines groups of the individuals
330
The processing system 10 determines diagnostic templates
340
The processing system 10 characterises the ability of the diagnostic templates to predict group membership
350
The templates are stored in the database
360

Fig. 6

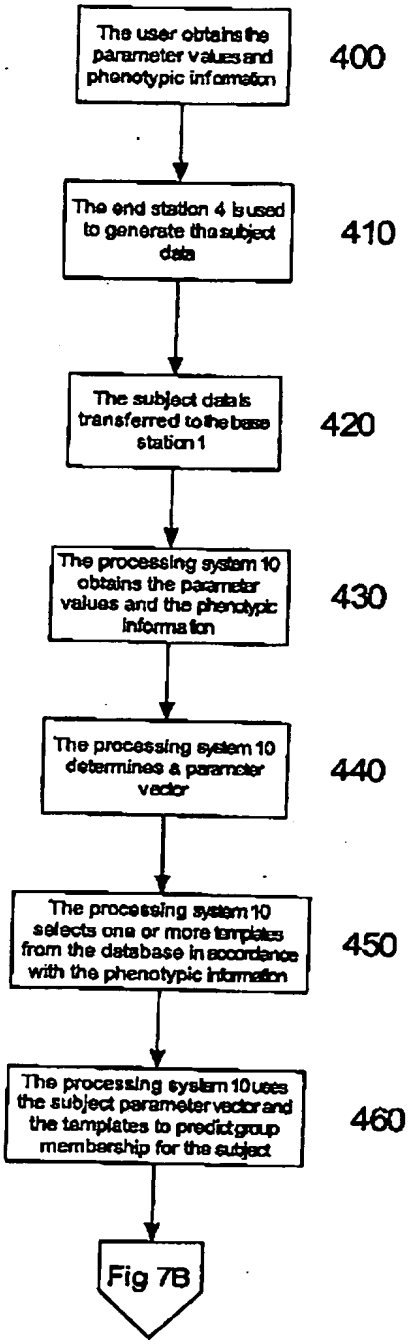
The user obtains the parameter values and phenotypic information
400
The end station 4 is used to generate the subject data
410
The subject data is transferred to the base station 1
420
The processing system 10 obtains the parameter values and the phenotypic information
430
The processing system 10 determines a parameter vector
440
The processing system 10 selects one or more templates from the database in accordance with the phenotypic information
450
The processing system 10 uses the subject parameter vector and the templates to predict group membership for the subject
460
Fig 7B

Fig. 7A

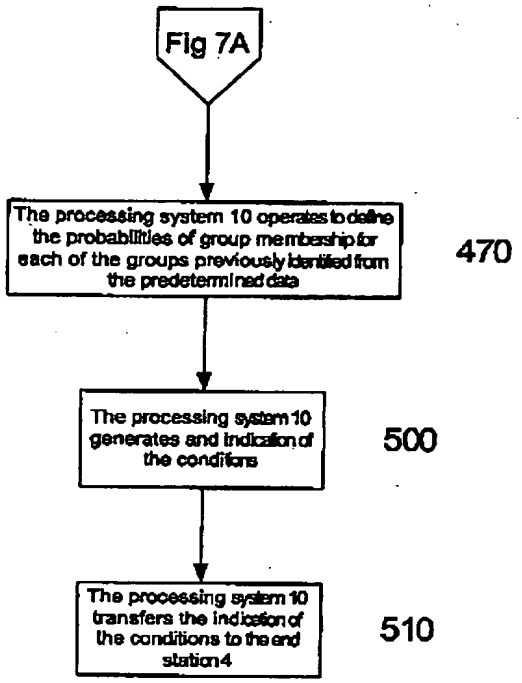
Fig 7A
The processing system 10 operates to define the probabilities of group membership for each of the groups previously identified from the predetermined data
470
The processing system 10 generates and indication of the conditions
500
The processing system 10 transfers the indication of the conditions to the end station4
510

Fig. 7B

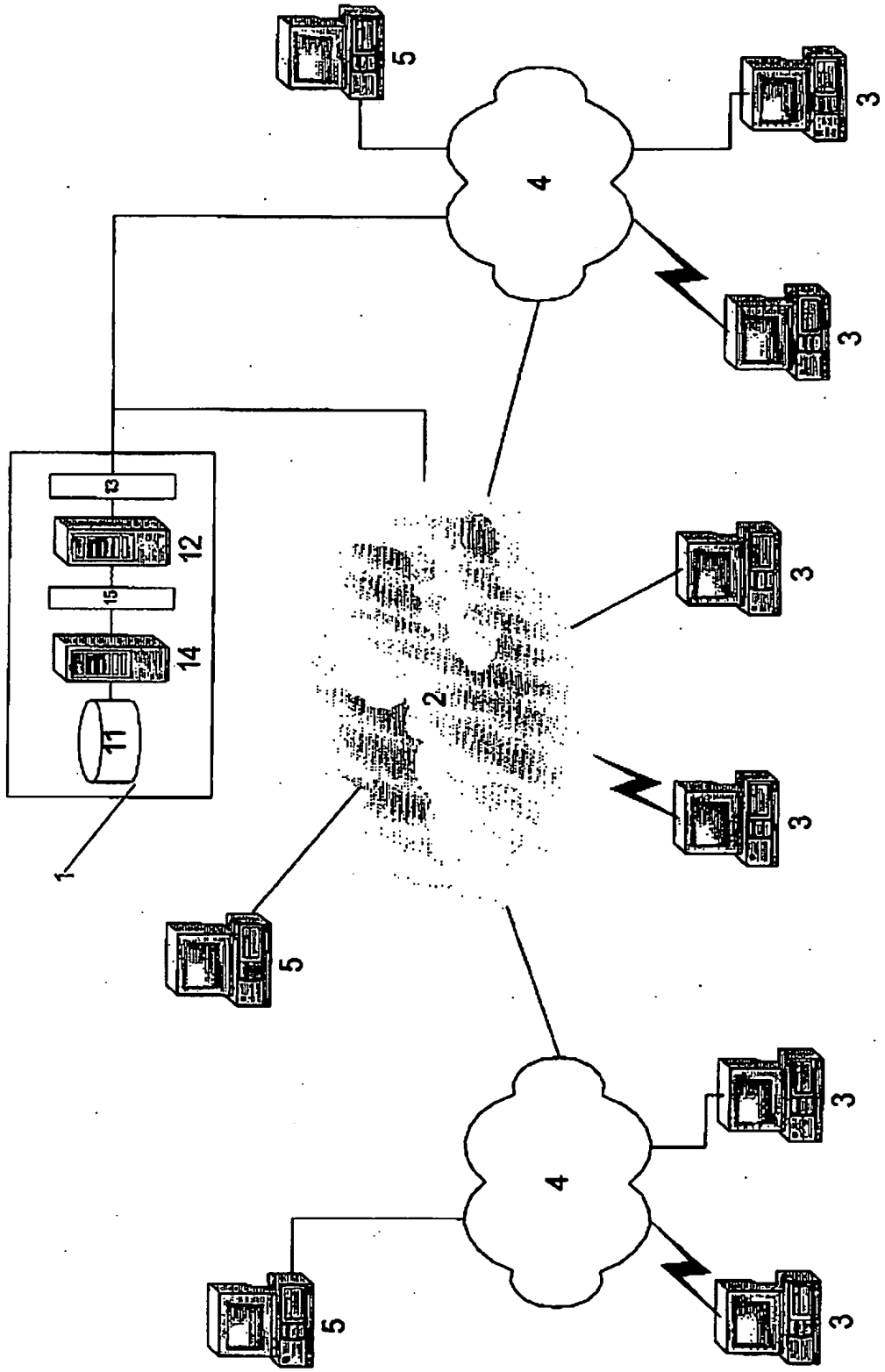

Fig. 8

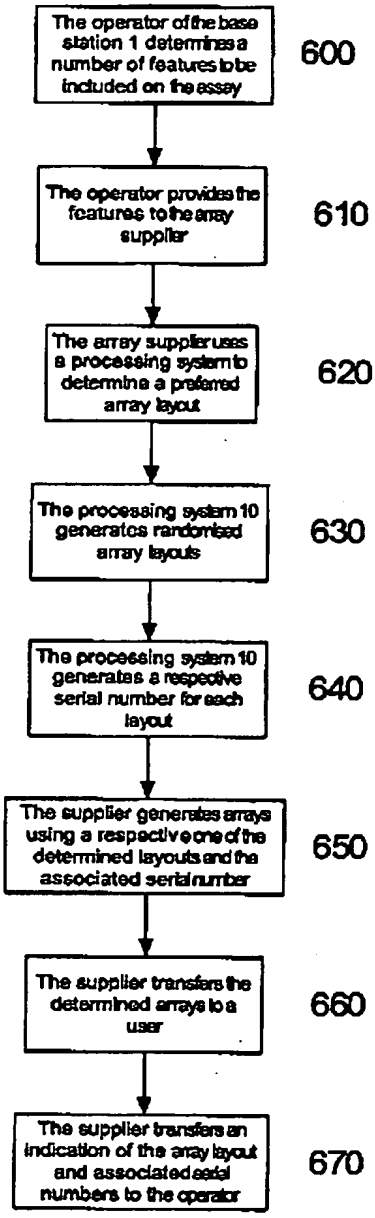
The operator of the base station 1 determines a number of features to be included on the assay
600
The operator provides the features to the array supplier
610
The array supplier uses a processing system to determine a preferred array layout
620
The processing system 10 generates randomised array layouts
630
The processing system 10 generates a respective serial number for each layout
640
The supplier generates arrays using a respective one of the determined layouts and the associated serial number
650
The supplier transfers the determined arrays to a user
660
The supplier transfers an indication of the array layout and associated serial numbers to the operator
670

Fig. 9

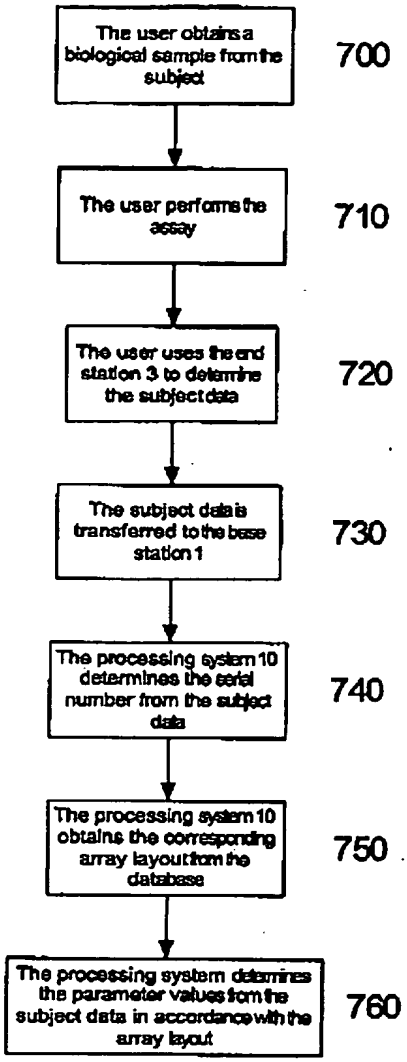
The user obtains a biological sample from the subject
700
The user performs the assay
710
The user uses the end station 3 to determine the subject data
720
The subject data is transferred to the base station 1
730
The processing system 10 determines the serial number from the subject data
740
The processing system 10 obtains the corresponding array layout from the database
750
The processing system determines the parameter values from the subject data in accordance with the array layout
760

Fig. 10

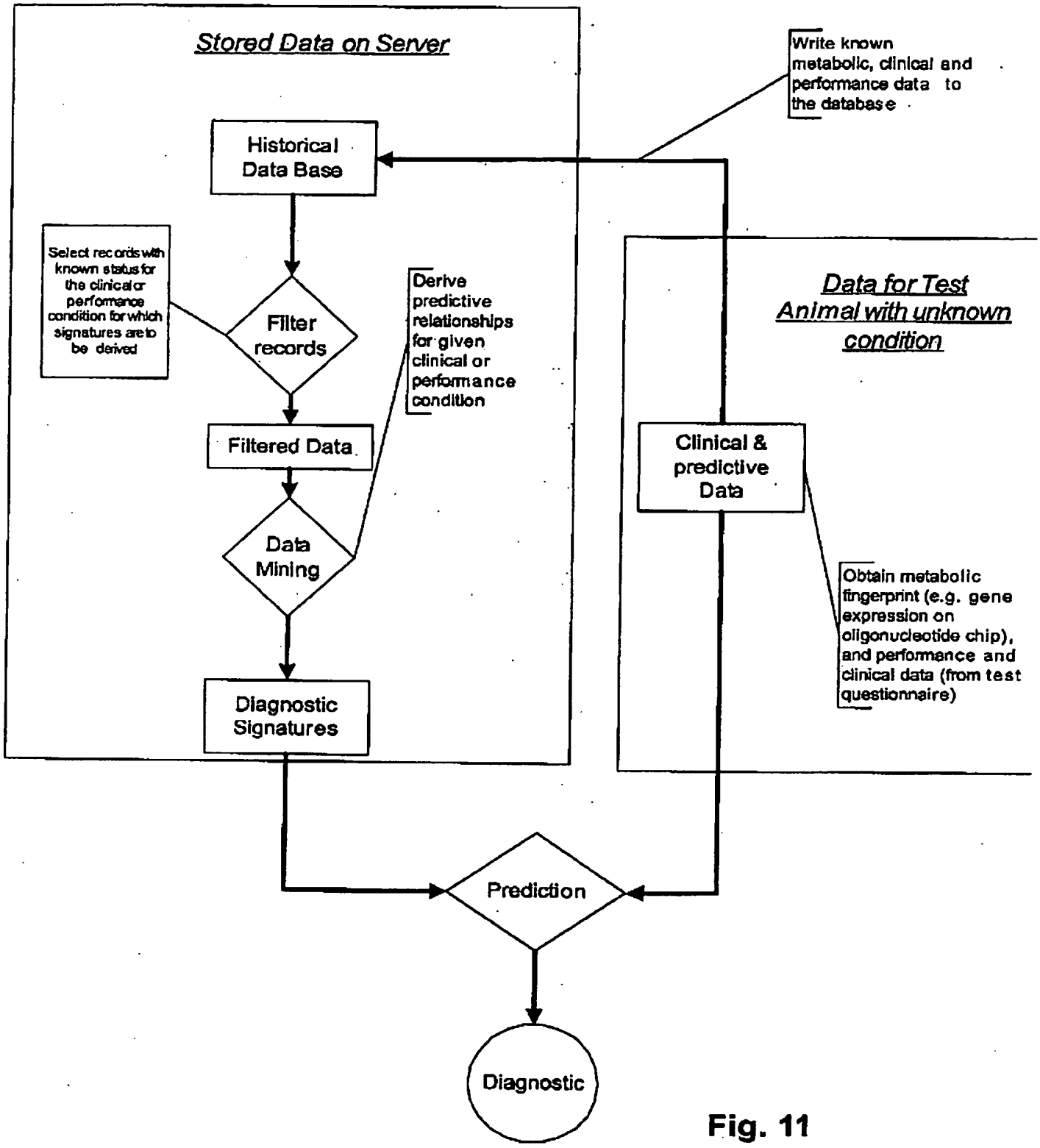
Stored Data on Server
Write known metabolic, clinical and performance data to the database
Historical Data Base
Select records with known status for the clinical or performance condition for which signatures are to be derived
Filter records
Derive predictive relationships for given clinical or performance condition
Filtered Data
Data Mining
Diagnostic Signatures
Data for Test Animal with unknown condition
Clinical & predictive Data
Obtain metabolic fingerprint (e.g. gene expression on oligonucleotide chip), and performance and clinical data (from test questionnaire)
Prediction
Diagnostic

Fig. 11

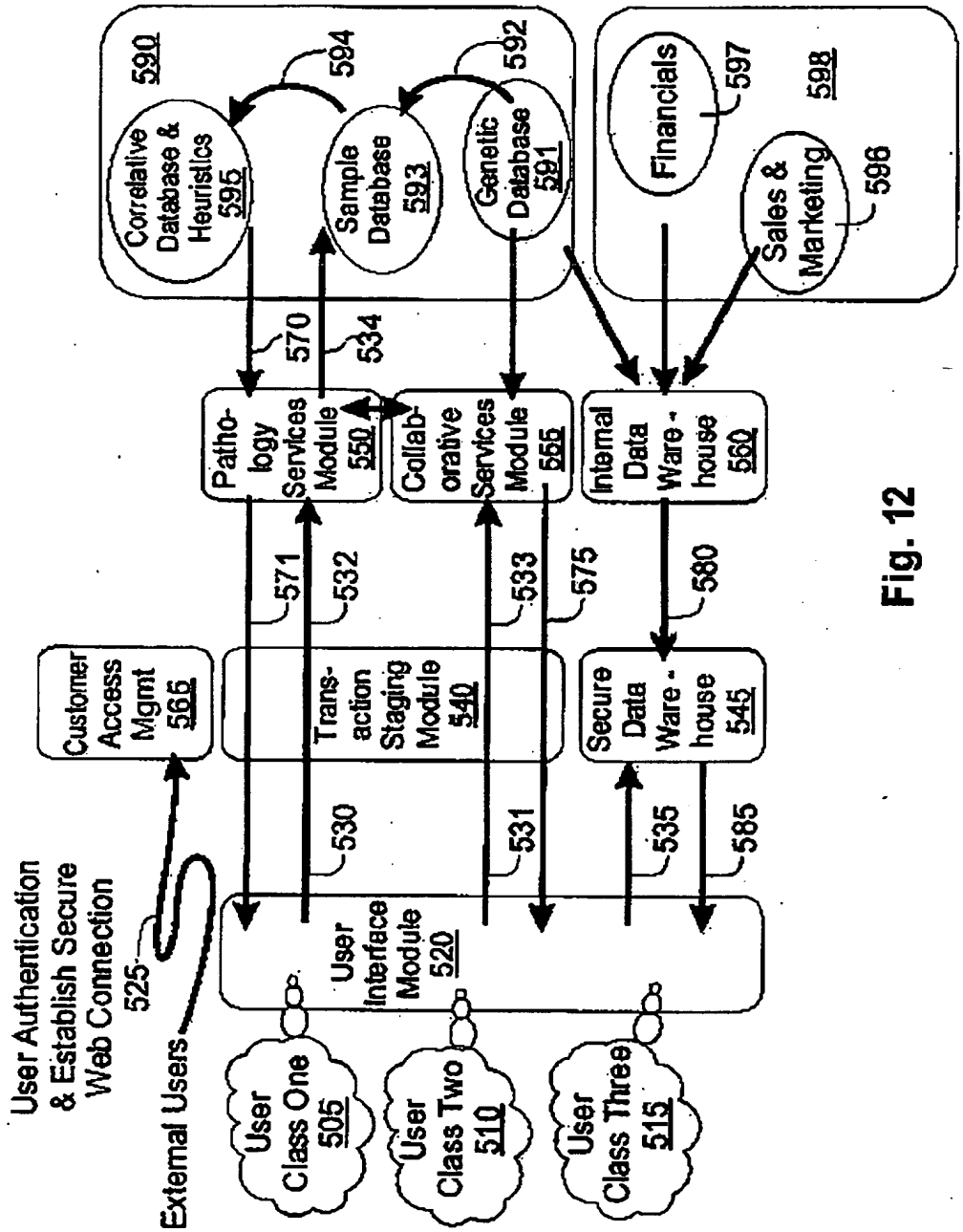
590
Correlative Database & Heuristics 595
594
Sample Database 593
592
Genetic Database 591
598
Financials 597
Sales & Marketing 596
570
534
Patho-logy Services Module 550
Collab-orative Services Module 555
Internal Data Ware-house 560
571
532
533
575
580
Customer Access Mgmt 566
Trans-action Staging Module 540
Secure Data Ware-house 545
530
531
535
585
User Interface Module 520
User Authentication & Establish Secure Web Connection 525
External Users
User Class One 505
User Class Two 510
User Class Three 515

Fig. 12

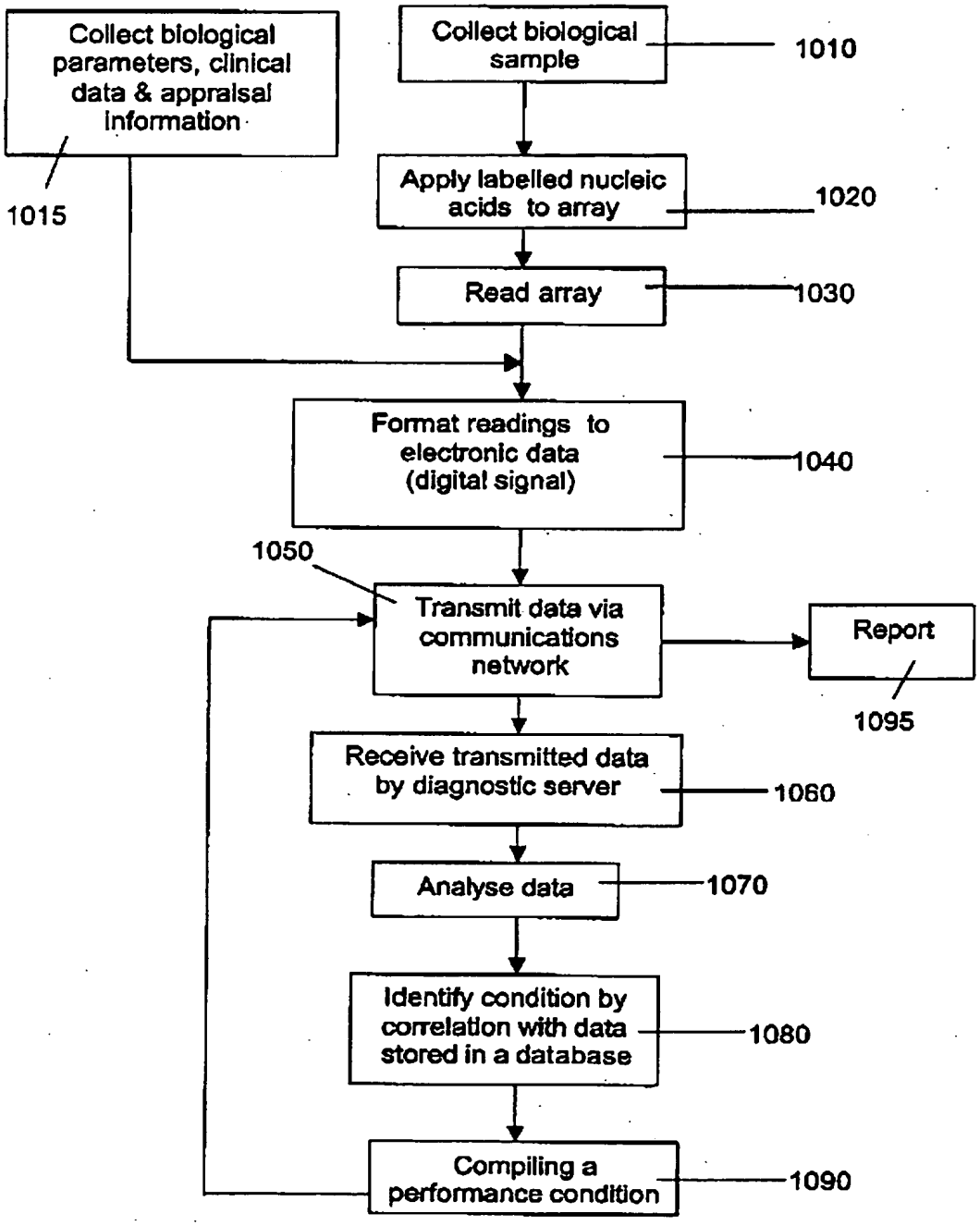
Collect biological parameters, clinical data & appraisal information
1015
Collect biological sample
1010
Apply labelled nucleic acids to array
1020
Read array
1030
Format readings to electronic data (digital signal)
1040
1050
Transmit data via communications network
Report
1095
Receive transmitted data by diagnostic server
1060
Analyse data
1070
Identify condition by correlation with data stored in a database
1080
Compiling a performance condition
1090

FIG. 13

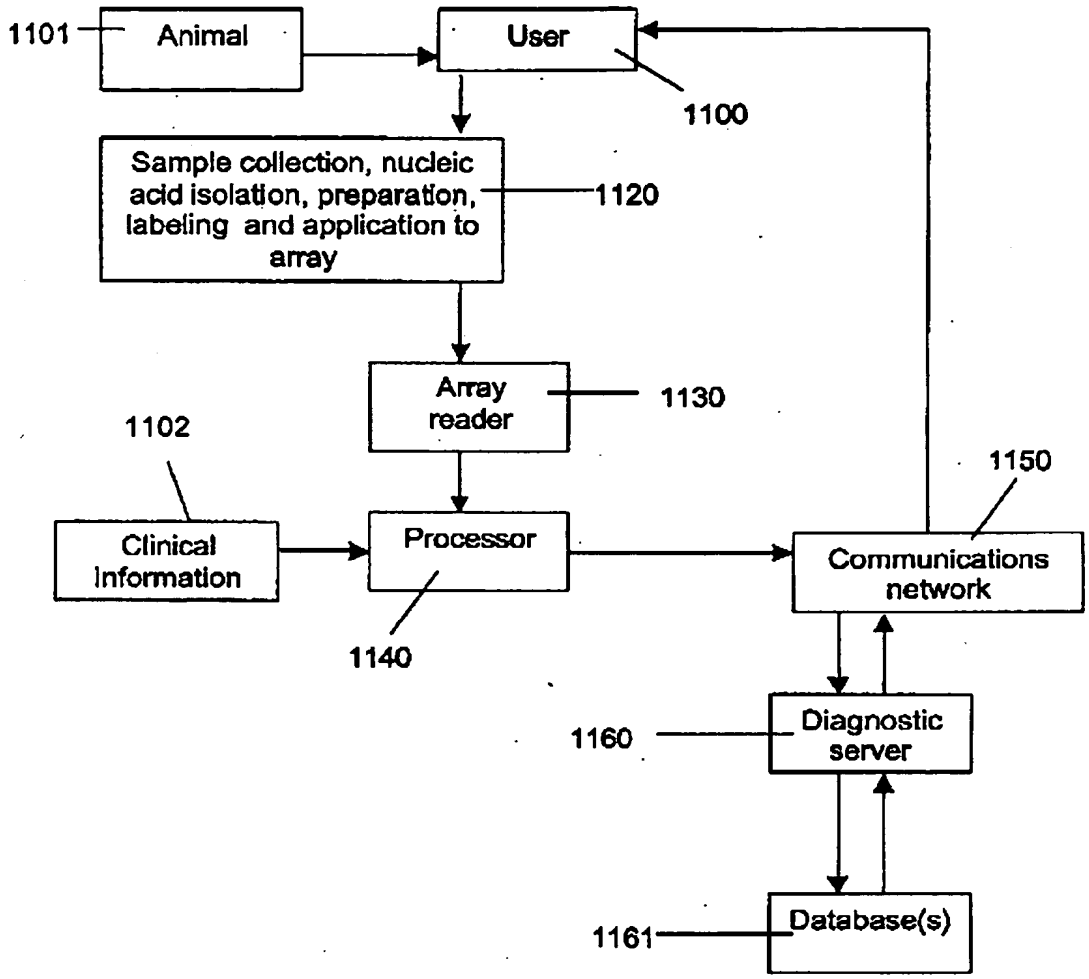
1101
Animal
User
1100
Sample collection, nucleic acid isolation, preparation, labeling and application to array
1120
Array reader
1130
1102
1150
Clinical Information
Processor
Communications network
1140
1160
Diagnostic server
1161
Database(s)

FIG. 14

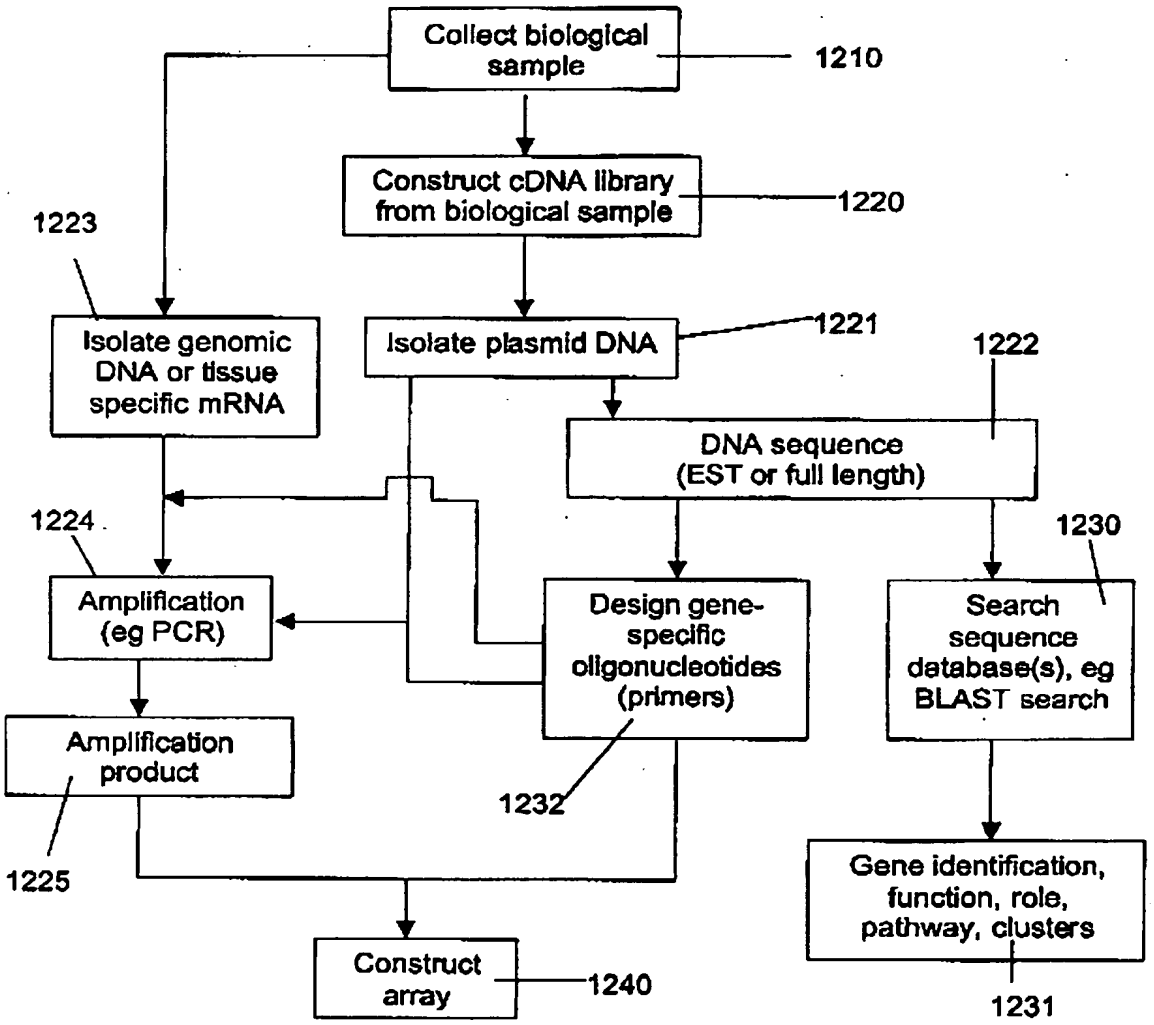
Collect biological sample
1210
Construct cDNA library from biological sample
1220
1223
Isolate genomic DNA or tissue specific mRNA
Isolate plasmid DNA
1221
1222
DNA sequence (EST or full length)
1224
1230
Amplification (eg PCR)
Design gene-specific oligonucleotides (primers)
Search sequence database(s), eg BLAST search
Amplification product
1232
1225
Gene identification, function, role, pathway, clusters
Construct array
1240
1231

FIG. 15

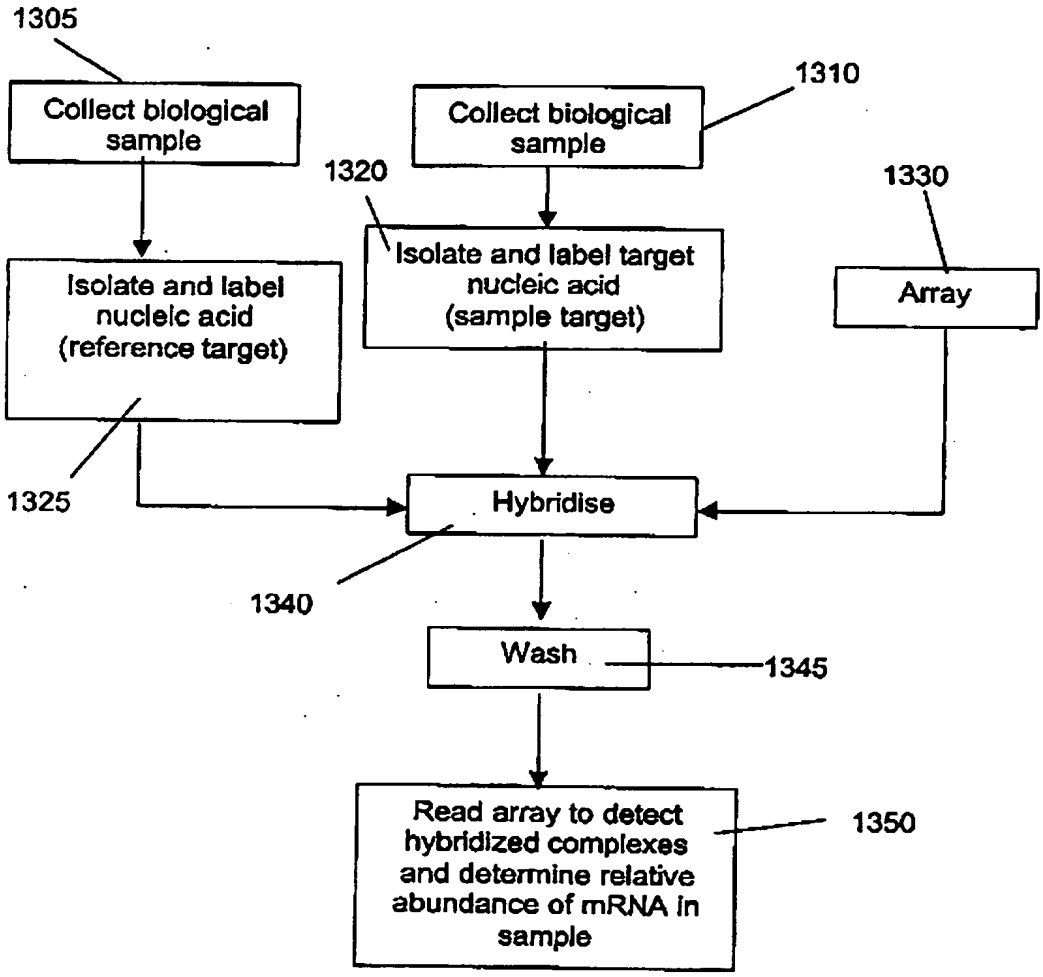
1305
Collect biological sample
1310
Collect biological sample
1320
1330
Isolate and label target nucleic acid (sample target)
Array
Isolate and label nucleic acid (reference target)
1325
Hybridise
1340
Wash
1345
Read array to detect hybridized complexes and determine relative abundance of mRNA in sample
1350

FIG. 16

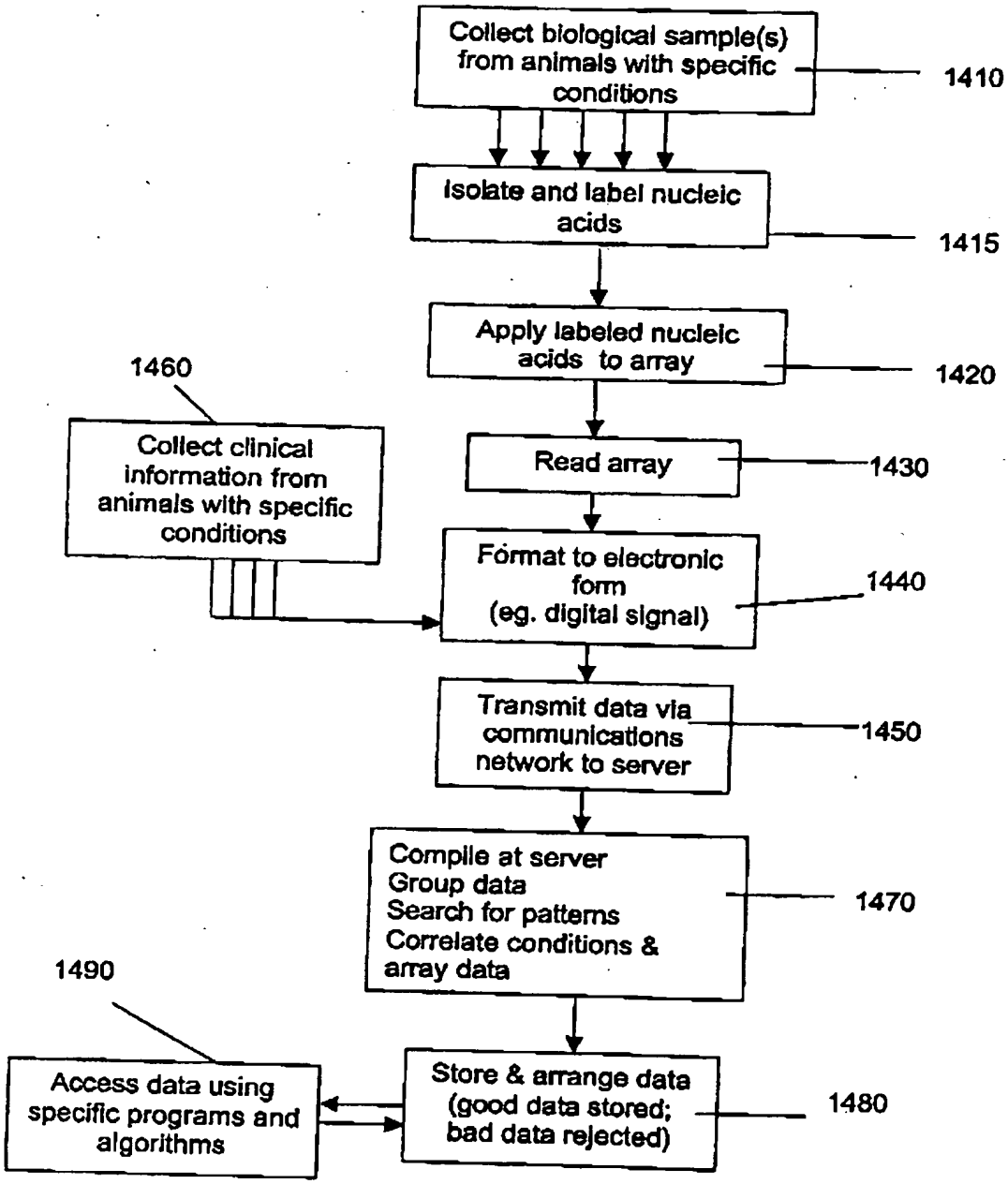
Collect biological sample(s) from animals with specific conditions
1410
Isolate and label nucleic acids
1415
Apply labeled nucleic acids to array
1420
1460
Collect clinical information from animals with specific conditions
Read array
1430
Format to electronic form (eg. digital signal)
1440
Transmit data via communications network to server
1450
Compile at server
Group data
Search for patterns
Correlate conditions & array data
1470
1490
Access data using specific programs and algorithms
Store & arrange data (good data stored; bad data rejected)
1480

Fig. 17